ADVANCED CALCULUS

A TEXT UPON SELECT PARTS OF DIFFERENTIAL CALCULUS, DIFFERENTIAL EQUATIONS, INTEGRAL CALCULUS, THEORY OF FUNCTIONS, WITH NUMEROUS EXERCISES

BY

EDWIN BIDWELL WILSON, Ph.D.

PROFESSOR OF MATHEMATICAL PHYSICS IN THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

GINN AND COMPANY

BOSTON · NEW YORK · CHICAGO · LONDON
ATLANTA · DALLAS · COLUMBUS · SAN FRANCISCO

121.6

The Athenæum Press
GINN AND COMPANY · PROPRIETORS · BOSTON · U.S.A.

PREFACE

It is probable that almost every teacher of advanced calculus feels the need of a text suited to present conditions and adaptable to his use. To write such a book is extremely difficult, for the attainments of students who enter a second course in calculus are different, their needs are not uniform, and the viewpoint of their teachers is no less varied. Yet in view of the cost of time and money involved in producing an Advanced Calculus, in proportion to the small number of students who will use it, it seems that few teachers can afford the luxury of having their own text; and that it consequently devolves upon an author to take as unselfish and unprejudiced a view of the subject as possible, and, so far as in him lies, to produce a book which shall have the maximum flexibility and adaptability. It was the recognition of this duty that has kept the present work in a perpetual state of growth and modification during five or six years of composition. Every attempt has been made to write in such a manner that the individual teacher may feel the minimum embarrassment in picking and choosing what seems to him best to meet the needs of any particular class.

As the aim of the book is to be a working text or laboratory manual for classroom use rather than an artistic treatise on analysis, especial attention has been given to the preparation of numerous exercises which should range all the way from those which require nothing but substitution in certain formulas to those which embody important results withheld from the text for the purpose of leaving the student some vital bits of mathematics to develop. It has been fully recognized that for the student of mathematics the work on advanced calculus falls in a period of transition, — of adolescence, — in which he must grow from close reliance upon his book to a large reliance upon himself. Moreover, as a course in advanced calculus is the *ultima Thule* of the mathematical voyages of most students of physics and engineering, it is appropriate that the text placed in the hands of those who seek that goal should by its method cultivate in them the attitude of courageous

explorers, and in its extent supply not only their immediate needs, but much that may be useful for later reference and independent study.

With the large necessities of the physicist and the growing requirements of the engineer, it is inevitable that the great majority of our students of calculus should need to use their mathematics readily and vigorously rather than with hesitation and rigor. Hence, although due attention has been paid to modern questions of rigor, the chief desire has been to confirm and to extend the student's working knowledge of those great algorisms of mathematics which are naturally associated with the calculus. That the compositor should have set "vigor" where "rigor" was written, might appear more amusing were it not for the suggested antithesis that there may be many who set rigor where vigor should be.

As I have had practically no assistance with either the manuscript or the proofs, I cannot expect that so large a work shall be free from errors; I can only have faith that such errors as occur may not prove seriously troublesome. To spend upon this book so much time and energy which could have been reserved with keener pleasure for various fields of research would have been too great a sacrifice, had it not been for the hope that I might accomplish something which should be of material assistance in solving one of the most difficult problems of mathematical instruction, — that of advanced calculus.

EDWIN BIDWELL WILSON

Massachusetts Institute of Technology

CONTENTS

INTRODUCTORY REVIEW

CHAPTER I

REVIEW OF FUNDAMENTAL RULES

CHAPTER II

REVIEW OF FUNDAMENTAL THEORY

PART I. DIFFERENTIAL CALCULUS

CHAPTER III

TAYLOR'S FORMULA AND ALLIED TOPICS

CHAPTER IV

PARTIAL DIFFERENTIATION; EXPLICIT FUNCTIONS

CHAPTER V

PARTIAL DIFFERENTIATION; IMPLICIT FUNCTIONS

CHAPTER VI

COMPLEX NUMBERS AND VECTORS

PART II. DIFFERENTIAL EQUATIONS

CHAPTER VII

GENERAL INTRODUCTION TO DIFFERENTIAL EQUATIONS

CHAPTER VIII

THE COMMONER ORDINARY DIFFERENTIAL EQUATIONS

CHAPTER IX

ADDITIONAL TYPES OF ORDINARY EQUATIONS

CHAPTER X

DIFFERENTIAL EQUATIONS IN MORE THAN TWO VARIABLES

PART III. INTEGRAL CALCULUS

CHAPTER XI

ON SIMPLE INTEGRALS

CHAPTER XII

ON MULTIPLE INTEGRALS

CHAPTER XIII

ON INFINITE INTEGRALS

CHAPTER XIV

SPECIAL FUNCTIONS DEFINED BY INTEGRALS

CHAPTER XV

THE CALCULUS OF VARIATIONS

PART IV. THEORY OF FUNCTIONS

CHAPTER XVI

INFINITE SERIES

CHAPTER XVII

SPECIAL INFINITE DEVELOPMENTS

CHAPTER XVIII

FUNCTIONS OF A COMPLEX VARIABLE

CHAPTER XIX

ELLIPTIC FUNCTIONS AND INTEGRALS

CHAPTER XX

FUNCTIONS OF REAL VARIABLES

ADVANCED CALCULUS

INTRODUCTORY REVIEW

CHAPTER I

REVIEW OF FUNDAMENTAL RULES

1. On differentiation. If the function $f(x)$ is interpreted as the curve $y = f(x)$,* the quotient of the increments Δy and Δx of the dependent and independent variables measured from (x_0, y_0) is

$$\frac{y - y_0}{x - x_0} = \frac{\Delta y}{\Delta x} = \frac{\Delta f(x)}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}, \tag{1}$$

and represents the *slope of the secant* through the points $P(x_0, y_0)$ and $P'(x_0 + \Delta x, y_0 + \Delta y)$ on the curve. The limit approached by the quotient $\Delta y / \Delta x$ when P remains fixed and $\Delta x \doteq 0$ is the *slope of the tangent* to the curve at the point P. *This limit,*

$$\lim_{\Delta x \doteq 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \doteq 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = f'(x_0), \tag{2}$$

is called the *derivative* of $f(x)$ for the value $x = x_0$. As the derivative may be computed for different points of the curve, it is customary to speak of the derivative as itself a function of x and write

$$\lim_{\Delta x \doteq 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \doteq 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x). \tag{3}$$

There are numerous notations for the derivative, for instance

$$f'(x) = \frac{df(x)}{dx} = \frac{dy}{dx} = D_x f = D_x y = y' = Df = Dy.$$

* Here and throughout the work, where figures are not given, the reader should draw graphs to illustrate the statements. Training in making one's own illustrations, whether graphical or analytic, is of great value.

The first five show distinctly that the independent variable is x, whereas the last three do not explicitly indicate the variable and should not be used unless there is no chance of a misunderstanding.

2. The fundamental formulas of differential calculus are derived directly from the application of the definition (2) or (3) and from a few fundamental propositions in limits. First may be mentioned

$$\frac{dz}{dx}=\frac{dz}{dy}\frac{dy}{dx}, \text{ where } z=\phi(y) \text{ and } y=f(x). \tag{4}$$

$$\frac{dx}{dy}=\frac{df^{-1}(y)}{dy}=\frac{1}{\dfrac{df(x)}{dx}}=\frac{1}{\dfrac{dy}{dx}}. \tag{5}$$

$$D(u\pm v)=Du\pm Dv, \qquad D(uv)=uDv+vDu. \tag{6}$$

$$D\left(\frac{u}{v}\right)=\frac{vDu-uDv}{v^2},^* \qquad D(x^n)=nx^{n-1}. \tag{7}$$

It may be recalled that (4), which is the rule for differentiating a function of a function, follows from the application of the theorem that the limit of a product is the product of the limits to the fractional identity $\frac{\Delta z}{\Delta x}=\frac{\Delta z}{\Delta y}\frac{\Delta y}{\Delta x}$; whence

$$\lim_{\Delta x\doteq 0}\frac{\Delta z}{\Delta x}=\lim_{\Delta x\doteq 0}\frac{\Delta z}{\Delta y}\cdot\lim_{\Delta x\doteq 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta y\doteq 0}{}^{\dagger}\frac{\Delta z}{\Delta y}\cdot\lim_{\Delta x\doteq 0}\frac{\Delta y}{\Delta x},$$

which is equivalent to (4). Similarly, if $y=f(x)$ and if x, as the inverse function of y, be written $x=f^{-1}(y)$ from analogy with $y=\sin x$ and $x=\sin^{-1}y$, the relation (5) follows from the fact that $\Delta x/\Delta y$ and $\Delta y/\Delta x$ are reciprocals. The next three result from the immediate application of the theorems concerning limits of sums, products, and quotients (§ 21). The rule for differentiating a power is derived in case n is integral by the application of the binomial theorem.

$$\frac{\Delta y}{\Delta x}=\frac{(x+\Delta x)^n-x^n}{\Delta x}=nx^{n-1}+\frac{n(n-1)}{2\,!}x^{n-2}\Delta x+\cdots+(\Delta x)^{n-1},$$

and the limit when $\Delta x\doteq 0$ is clearly nx^{n-1}. The result may be extended to rational values of the index n by writing $n=\frac{p}{q}$, $y=x^{\frac{p}{q}}$, $y^q=x^p$ and by differentiating both sides of the equation and reducing. To prove that (7) still holds when n is irrational, it would be *necessary to have a workable definition of irrational numbers* and to develop the properties of such numbers in greater detail than seems wise at this point. The formula is therefore assumed in accordance with *the principle of permanence of form* (§ 178), just as formulas like $a^m a^n=a^{m+n}$ of the theory of exponents, which may readily be proved for rational bases and exponents, are assumed without proof to hold also for irrational bases and exponents. See, however, §§ 18–25 and the exercises thereunder.

* It is frequently better to regard the quotient as the product $u\cdot v^{-1}$ and apply (6).

† For when $\Delta x \doteq 0$, then $\Delta y\doteq 0$ or $\Delta y/\Delta x$ could not approach a limit.

3. Second may be mentioned the formulas for the derivatives of the trigonometric and the inverse trigonometric functions.

$$D\sin x = \cos x, \qquad D\cos x = -\sin x, \tag{8}$$

or

$$D\sin x = \sin\left(x + \tfrac{1}{2}\pi\right), \qquad D\cos x = \cos\left(x + \tfrac{1}{2}\pi\right), \tag{8'}$$

$$D\tan x = \sec^2 x, \qquad D\cot x = -\csc^2 x, \tag{9}$$

$$D\sec x = \sec x\tan x, \qquad D\csc x = -\csc x\cot x, \tag{10}$$

$$D\operatorname{vers} x = \sin x, \quad \text{where} \quad \operatorname{vers} x = 1 - \cos x = 2\sin^2\tfrac{1}{2}x, \tag{11}$$

$$D\sin^{-1}x = \frac{\pm 1}{\sqrt{1-x^2}}, \qquad \begin{cases} + \text{ in quadrants I, IV,} \\ - \text{ “ “ II, III,} \end{cases} \tag{12}$$

$$D\cos^{-1}x = \frac{\pm 1}{\sqrt{1-x^2}}, \qquad \begin{cases} - \text{ in quadrants I, II,} \\ + \text{ “ “ III, IV,} \end{cases} \tag{13}$$

$$D\tan^{-1}x = \frac{1}{1+x^2}, \qquad D\cot^{-1}x = -\frac{1}{1+x^2}, \tag{14}$$

$$D\sec^{-1}x = \frac{\pm 1}{x\sqrt{x^2-1}}, \qquad \begin{cases} + \text{ in quadrants I, III,} \\ - \text{ “ “ II, IV,} \end{cases} \tag{15}$$

$$D\csc^{-1}x = \frac{\pm 1}{x\sqrt{x^2-1}}, \qquad \begin{cases} - \text{ in quadrants I, III,} \\ + \text{ “ “ II, IV,} \end{cases} \tag{16}$$

$$D\operatorname{vers}^{-1}x = \frac{\pm 1}{\sqrt{2x-x^2}}, \qquad \begin{cases} + \text{ in quadrants I, II,} \\ - \text{ “ “ III, IV.} \end{cases} \tag{17}$$

It may be recalled that to differentiate $\sin x$ the definition is applied. Then

$$\frac{\Delta\sin x}{\Delta x} = \frac{\sin(x+\Delta x) - \sin x}{\Delta x} = \frac{\sin\Delta x}{\Delta x}\cos x - \frac{1-\cos\Delta x}{\Delta x}\sin x.$$

It now is merely a question of evaluating the two limits which thus arise, namely,

$$\lim_{\Delta x \doteq 0}\frac{\sin\Delta x}{\Delta x} \quad \text{and} \quad \lim_{\Delta x \doteq 0}\frac{1-\cos\Delta x}{\Delta x}. \tag{18}$$

From the properties of the circle it follows that these are respectively 1 and 0. Hence the derivative of $\sin x$ is $\cos x$. The derivative of $\cos x$ may be found in like manner or from the identity $\cos x = \sin\left(\frac{1}{2}\pi - x\right)$. The results for all the other trigonometric functions are derived by expressing the functions in terms of $\sin x$ and $\cos x$. And to treat the inverse functions, it is sufficient to recall the general method in (5). Thus

$$\text{if} \quad y = \sin^{-1}x, \qquad \text{then} \quad \sin y = x.$$

Differentiate both sides of the latter equation and note that $\cos y = \pm\sqrt{1-\sin^2 y} = \pm\sqrt{1-x^2}$ and the result for $D\sin^{-1}x$ is immediate. To ascertain which sign to use with the radical, it is sufficient to note that $\pm\sqrt{1-x^2}$ is $\cos y$, which is positive when the angle $y = \sin^{-1}x$ is in quadrants I and IV, negative in II and III. Similarly for the other inverse functions.

EXERCISES *

1. Carry through the derivation of (7) when $n = p/q$, and review the proofs of typical formulas selected from the list (5)–(17). Note that the formulas are often given as $D_x u^n = nu^{n-1} D_x u$, $D_x \sin u = \cos u \, D_x u$, $\cdots$, and may be derived in this form directly from the definition (3).

2. Derive the two limits necessary for the differentiation of $\sin x$.

3. Draw graphs of the inverse trigonometric functions and label the portions of the curves which correspond to quadrants I, II, III, IV. Verify the sign in (12)–(17) from the slope of the curves.

4. Find $D \tan x$ and $D \cot x$ by applying the definition (3) directly.

5. Find $D \sin x$ by the identity $\sin u - \sin v = 2 \cos \dfrac{u+v}{2} \sin \dfrac{u-v}{2}$.

6. Find $D \tan^{-1} x$ by the identity $\tan^{-1} u - \tan^{-1} v = \tan^{-1} \dfrac{u-v}{1+uv}$ and (3).

7. Differentiate the following expressions:

(α) $\csc 2x - \cot 2x$, (β) $\frac{1}{3} \tan^3 x - \tan x + x$, (γ) $x \cos^{-1} x - \sqrt{1 - x^2}$,

(δ) $\sec^{-1} \dfrac{1}{\sqrt{1-x^2}}$, (ϵ) $\sin^{-1} \dfrac{x}{\sqrt{1+x^2}}$, (ζ) $x\sqrt{a^2 - x^2} + a^2 \sin^{-1} \dfrac{x}{a}$,

(η) $a \operatorname{vers}^{-1} \dfrac{x}{a} - \sqrt{2ax - x^2}$, (θ) $\cot^{-1} \dfrac{2ax}{x^2 - a^2} - 2 \tan^{-1} \dfrac{x}{a}$.

What trigonometric identities are suggested by the answers for the following:

(α) $\sec^2 x$, (δ) $\dfrac{1}{\sqrt{1-x^2}}$, (ϵ) $\dfrac{1}{1+x^2}$, (θ) 0?

8. In B. O. Peirce's "Short Table of Integrals" (revised edition) differentiate the right-hand members to confirm the formulas: Nos. 31, 45–47, 91–97, 125, 127–128, 131–135, 161–163, 214–216, 220, 260–269, 294–298, 300, 380–381, 386–394.

9. If x is measured in degrees, what is $D \sin x$?

4. The logarithmic, exponential, and hyperbolic functions. The next set of formulas to be cited are

$$D \log_e x = \frac{1}{x}, \qquad D \log_a x = \frac{\log_a e}{x}, \tag{19}$$

$$De^x = e^x, \qquad Da^x = a^x \log_e a.\dagger \tag{20}$$

It may be recalled that the procedure for differentiating the logarithm is

$$\frac{\Delta \log_a x}{\Delta x} = \frac{\log_a (x + \Delta x) - \log_a x}{\Delta x} = \frac{1}{\Delta x} \log_a \frac{x + \Delta x}{x} = \frac{1}{x} \log_a \left(1 + \frac{\Delta x}{x}\right)^{\frac{x}{\Delta x}}.$$

* The student should keep on file his solutions of at least the important exercises; many subsequent exercises and considerable portions of the text depend on previous exercises.

† As is customary, the subscript e will hereafter be omitted and the symbol log will denote the logarithm to the base e; any base other than e must be specially designated as such. This observation is particularly necessary with reference to the common base 10 used in computation.

If now $x/\Delta x$ be set equal to h, the problem becomes that of evaluating

$$\lim_{h=\infty}\left(1+\frac{1}{h}\right)^h = e = 2.71828\cdots,^* \qquad \log_{10} e = 0.434294\cdots; \tag{21}$$

and hence if e be chosen as the base of the system, $D\log x$ takes the simple form $1/x$. The exponential functions e^x and a^x may be regarded as the inverse functions of $\log x$ and $\log_a x$ in deducing (21). Further it should be noted that it is frequently useful to take the logarithm of an expression before differentiating. This is known as *logarithmic differentiation* and is used for products and complicated powers and roots. Thus

$$\text{if} \qquad y = x^x, \qquad \text{then} \qquad \log y = x\log x,$$

$$\text{and} \qquad \frac{1}{y}y' = 1 + \log x \qquad \text{or} \qquad y' = x^x(1+\log x).$$

It is the expression y'/y which is called the *logarithmic derivative* of y. An especially noteworthy property of the function $y = Ce^x$ is that the function and its derivative are equal, $y' = y$; and more generally *the function $y = Ce^{kx}$ is proportional to its derivative*, $y' = ky$.

5. The *hyperbolic functions* are the hyperbolic sine and cosine,

$$\sinh x = \frac{e^x - e^{-x}}{2}, \qquad \cosh x = \frac{e^x + e^{-x}}{2}; \tag{22}$$

and the related functions $\tanh x$, $\coth x$, $\operatorname{sech} x$, $\operatorname{csch} x$, derived from them by the same ratios as those by which the corresponding trigonometric functions are derived from $\sin x$ and $\cos x$. From these definitions in terms of exponentials follow the formulas:

$$\cosh^2 x - \sinh^2 x = 1, \qquad \tanh^2 x + \operatorname{sech}^2 x = 1, \tag{23}$$

$$\sinh(x \pm y) = \sinh x \cosh y \pm \cosh x \sinh y, \tag{24}$$

$$\cosh(x \pm y) = \cosh x \cosh y \pm \sinh x \sinh y, \tag{25}$$

$$\cosh\frac{x}{2} = +\sqrt{\frac{\cosh x + 1}{2}}, \qquad \sinh\frac{x}{2} = \pm\sqrt{\frac{\cosh x - 1}{2}}, \tag{26}$$

$$D\sinh x = \cosh x, \qquad D\cosh x = \sinh x, \tag{27}$$

$$D\tanh x = \operatorname{sech}^2 x, \qquad D\coth x = -\operatorname{csch}^2 x, \tag{28}$$

$$D\operatorname{sech} x = -\operatorname{sech} x \tanh x, \qquad D\operatorname{csch} x = -\operatorname{csch} x \coth x. \tag{29}$$

The inverse functions are expressible in terms of logarithms. Thus

$$y = \sinh^{-1} x, \qquad x = \sinh y = \frac{e^{2y} - 1}{2e^y},$$

$$e^{2y} - 2xe^y - 1 = 0, \qquad e^y = x \pm \sqrt{x^2 + 1}.$$

* The treatment of this limit is far from complete in the majority of texts. Reference for a careful presentation may, however, be made to Granville's "Calculus," pp. 31–34, and Osgood's "Calculus," pp. 78–82. See also Ex. 1, (β), in § 165 below.

Here only the positive sign is available, for e^y is never negative. Hence

$$\sinh^{-1} x = \log\left(x + \sqrt{x^2+1}\right), \qquad \text{any } x, \tag{30}$$

$$\cosh^{-1} x = \log\left(x \pm \sqrt{x^2-1}\right), \qquad x > 1, \tag{31}$$

$$\tanh^{-1} x = \frac{1}{2}\log\frac{1+x}{1-x}, \qquad x^2 < 1, \tag{32}$$

$$\coth^{-1} x = \frac{1}{2}\log\frac{x+1}{x-1}, \qquad x^2 > 1, \tag{33}$$

$$\operatorname{sech}^{-1} x = \log\left(\frac{1}{x} \pm \sqrt{\frac{1}{x^2} - 1}\right), \qquad x < 1, \tag{34}$$

$$\operatorname{csch}^{-1} x = \log\left(\frac{1}{x} + \sqrt{\frac{1}{x^2} + 1}\right), \qquad \text{any } x, \tag{35}$$

$$D\sinh^{-1} x = \frac{+1}{\sqrt{x^2+1}}, \qquad D\cosh^{-1} x = \frac{\pm 1}{\sqrt{x^2-1}}, \tag{36}$$

$$D\tanh^{-1} x = \frac{1}{1-x^2} = D\coth^{-1} x = \frac{1}{1-x^2}, \tag{37}$$

$$D\operatorname{sech}^{-1} x = \frac{\pm 1}{x\sqrt{1-x^2}}, \qquad D\operatorname{csch}^{-1} x = \frac{-1}{x\sqrt{1+x^2}}. \tag{38}$$

EXERCISES

1. Show by logarithmic differentiation that

$$D(uvw\cdots) = \left(\frac{u'}{u} + \frac{v'}{v} + \frac{w'}{w} + \cdots\right)(uvw\cdots),$$

and hence derive the rule: To differentiate a product differentiate each factor alone and add all the results thus obtained.

2. Sketch the graphs of the hyperbolic functions, interpret the graphs as those of the inverse functions, and verify the range of values assigned to x in (30)–(35).

3. Prove sundry of formulas (23)–(29) from the definitions (22).

4. Prove sundry of (30)–(38), checking the signs with care. In cases where double signs remain, state when each applies. Note that in (31) and (34) *the double sign may be placed before the* log *for the reason that the two expressions are reciprocals.*

5. Derive a formula for $\sinh u \pm \sinh v$ by applying (24); find a formula for $\tanh \frac{1}{2}x$ analogous to the trigonometric formula $\tan\frac{1}{2}x = \sin x/(1+\cos x)$.

6. *The gudermannian.* The function $\phi = \operatorname{gd} x$, defined by the relations

$$\sinh x = \tan\phi, \quad \phi = \operatorname{gd} x = \tan^{-1}\sinh x, \quad -\tfrac{1}{2}\pi < \phi < +\tfrac{1}{2}\pi,$$

is called the gudermannian of x. Prove the set of formulas:

$$\cosh x = \sec\phi, \quad \tanh x = \sin\phi, \quad \operatorname{csch} x = \cot\phi, \quad \text{etc.};$$

$$D\operatorname{gd} x = \operatorname{sech} x, \quad x = \operatorname{gd}^{-1}\phi = \log\tan\left(\tfrac{1}{2}\phi + \tfrac{1}{4}\pi\right), \quad D\operatorname{gd}^{-1}\phi = \sec\phi.$$

7. Substitute the functions of ϕ in Ex. 6 for their hyperbolic equivalents in (23), (26), (27), and reduce to simple known trigonometric formulas.

8. Differentiate the following expressions:

(α) $(x+1)^2(x+2)^{-3}(x+3)^{-4}$, (β) $x^{\log x}$, (γ) $\log_x(x+1)$,
(δ) $x+\log\cos(x-\frac{1}{4}\pi)$, (ϵ) $2\tan^{-1}e^x$, (ζ) $x-\tanh x$,
(η) $x\tanh^{-1}x+\frac{1}{2}\log(1-x^2)$, (θ) $\dfrac{e^{ax}(a\sin mx-m\cos mx)}{m^2+a^2}$.

9. Check sundry formulas of Peirce's "Table," pp. 1–61, 81–82.

6. Geometric properties of the derivative. As the quotient (1) and its limit (2) give the slope of a secant and of the tangent, it appears from graphical considerations that when the derivative is positive the function is increasing with x, but decreasing when the derivative is negative.* Hence *to determine the regions in which a function is increasing or decreasing, one may find the derivative and determine the regions in which it is positive or negative.*

One must, however, be careful not to apply this rule too blindly; for in so simple a case as $f(x)=\log x$ it is seen that $f'(x)=1/x$ is positive when $x>0$ and negative when $x<0$, and yet $\log x$ has no graph when $x<0$ and is not considered as decreasing. Thus the formal derivative may be real when the function is not real, and it is therefore best to make a rough sketch of the function to corroborate the evidence furnished by the examination of $f'(x)$.

If x_0 is a value of x such that immediately† upon one side of $x=x_0$ the function $f(x)$ is increasing whereas immediately upon the other side it is decreasing, the ordinate $y_0=f(x_0)$ will be a maximum or minimum or $f(x)$ will become positively or negatively infinite at x_0. If the case where $f(x)$ becomes infinite be ruled out, one may say that *the function will have a minimum or maximum at x_0 according as the derivative changes from negative to positive or from positive to negative when x, moving in the positive direction, passes through the value x_0.* Hence *the usual rule for determining maxima and minima is to find the roots of $f'(x)=0$.*

This rule, again, must not be applied blindly. For first, $f'(x)$ may vanish where there is no maximum or minimum as in the case $y=x^3$ at $x=0$ where the derivative does not change sign; or second, $f'(x)$ may change sign by becoming infinite as in the case $y=x^{\frac{2}{3}}$ at $x=0$ where the curve has a vertical cusp, point down, and a minimum; or third, the function $f(x)$ may be restricted to a given range of values $a\leqq x\leqq b$ for x and then the values $f(a)$ and $f(b)$ of the function at the ends of the interval will in general be maxima or minima without implying that the derivative vanish. Thus although the derivative is highly useful in determining maxima and minima, it should not be trusted to the complete exclusion of the corroborative evidence furnished by a rough sketch of the curve $y=f(x)$.

* The construction of illustrative figures is again left to the reader.

† The word "immediately" is necessary because the maxima or minima may be merely *relative*; in the case of several maxima and minima in an interval, some of the maxima may actually be less than some of the minima.

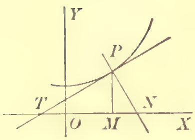
7. The derivative may be used to express the *equations of the tangent and normal*, the *values of the subtangent and subnormal*, and so on.

Equation of tangent, $\quad y - y_0 = y_0'(x - x_0)$, $\qquad$ (39)

Equation of normal, $\quad (y - y_0)\,y_0' + (x - x_0) = 0$, $\qquad$ (40)

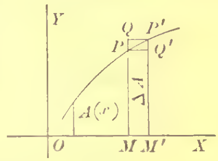
$$TM = \text{subtangent} = y_0/y_0', \quad MN = \text{subnormal} = y_0 y_0', \tag{41}$$

$$OT = x\text{-intercept of tangent} = x_0 - y_0/y_0', \text{ etc.} \tag{42}$$



The derivation of these results is sufficiently evident from the figure. It may be noted that the subtangent, subnormal, etc., are numerical values for a given point of the curve but may be regarded as functions of x like the derivative.

In geometrical and physical problems it is frequently necessary to apply the definition of the derivative to finding the derivative of an unknown function. For instance if A denote the area under a curve and measured from a fixed ordinate to a variable ordinate, A is surely a function $A(x)$ of the abscissa x of the variable ordinate. If the curve is rising, as in the figure, then



$$MPQ'M' < \Delta A < MQP'M', \text{ or } y\Delta x < \Delta A < (y + \Delta y)\Delta x.$$

Divide by Δx and take the limit when $\Delta x \doteq 0$. There results

$$\lim_{\Delta x \doteq 0} y \leqq \lim_{\Delta x \doteq 0} \frac{\Delta A}{\Delta x} \leqq \lim_{\Delta x \doteq 0} (y + \Delta y).$$

Hence
$$\lim_{\Delta x \doteq 0} \frac{\Delta A}{\Delta x} = \frac{dA}{dx} = y. \tag{43}$$

Rolle's Theorem and the *Theorem of the Mean* are two important theorems on derivatives which will be treated in the next chapter but may here be stated as evident from their geometric interpretation. Rolle's Theorem states that: *If a function has a derivative at every*

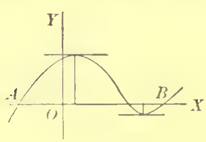


Fig. 1

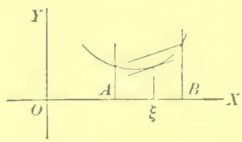


Fig. 2

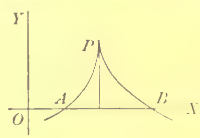


Fig. 3

point of an interval and if the function vanishes at the ends of the interval, then there is at least one point within the interval at which the derivative vanishes. This is illustrated in Fig. 1, in which there are two such points. The Theorem of the Mean states that: *If a function*

has a derivative at each point of an interval, there is at least one point in the interval such that the tangent to the curve $y = f(x)$ is parallel to the chord of the interval. This is illustrated in Fig. 2 in which there is only one such point.

Again care must be exercised. In Fig. 3 the function vanishes at A and B but there is no point at which the slope of the tangent is zero. This is not an exception or contradiction to Rolle's Theorem for the reason that the function does not satisfy the conditions of the theorem. In fact at the point P, although there is a tangent to the curve, there is no derivative; the quotient (1) formed for the point P becomes negatively infinite as $\Delta x \doteq 0$ from one side, positively infinite as $\Delta x \doteq 0$ from the other side, and therefore does not approach a definite limit as is required in the definition of a derivative. The hypothesis of the theorem is not satisfied and there is no reason that the conclusion should hold.

EXERCISES

1. Determine the regions in which the following functions are increasing or decreasing, sketch the graphs, and find the maxima and minima:

(α) $\frac{1}{3}x^3 - x^2 + 2$, (β) $(x+1)^{\frac{2}{3}}(x-5)^2$, (γ) $\log(x^2 - 4)$,
(δ) $(x-2)\sqrt{x-1}$, (ϵ) $-(x+2)\sqrt{12 - x^2}$, (ζ) $x^3 + ax + b$.

2. The ellipse is $r = \sqrt{x^2 + y^2} = e(d + x)$ referred to an origin at the focus. Find the maxima and minima of the focal radius r, and state why $D_x r = 0$ does not give the solutions while $D_\phi r = 0$ does [the polar form of the ellipse being $r = k(1 - e\cos\phi)^{-1}$].

3. Take the ellipse as $x^2/a^2 + y^2/b^2 = 1$ and discuss the maxima and minima of the central radius $r = \sqrt{x^2 + y^2}$. Why does $D_x r = 0$ give half the result when r is expressed as a function of x, and why will $D_\lambda r = 0$ give the whole result when $x = a\cos\lambda$, $y = b\sin\lambda$ and the ellipse is thus expressed in terms of the eccentric angle?

4. If $y = P(x)$ is a polynomial in x such that the equation $P(x) = 0$ has multiple roots, show that $P'(x) = 0$ for each multiple root. What more complete relationship can be stated and proved?

5. Show that the triple relation $27b^2 + 4a^3 \gtreqless 0$ determines completely the nature of the roots of $x^3 + ax + b = 0$, and state what corresponds to each possibility.

6. Define *the angle θ between two intersecting curves*. Show that

$$\tan\theta = [f'(x_0) - g'(x_0)] \div [1 + f'(x_0)g'(x_0)]$$

if $y = f(x)$ and $y = g(x)$ cut at the point (x_0, y_0).

7. Find the subnormal and subtangent of the three curves

(α) $y^2 = 4px$, (β) $x^2 = 4py$, (γ) $x^2 + y^2 = a^2$.

8. *The pedal curve.* The locus of the foot of the perpendicular dropped from a fixed point to a variable tangent of a given curve is called the pedal of the given curve with respect to the given point. Show that if the fixed point is the origin, the pedal of $y = f(x)$ may be obtained by eliminating x_0, y_0, y_0' from the equations

$$y - y_0 = y_0'(x - x_0), \quad yy_0' + x = 0, \quad y_0 = f(x_0), \quad y_0' = f'(x_0).$$

Find the pedal (α) of the hyperbola with respect to the center and (β) of the parabola with respect to the vertex and (γ) the focus. Show (δ) that the pedal of the parabola with respect to any point is a cubic.

9. If the curve $y = f(x)$ be revolved about the x-axis and if $V(x)$ denote the volume of revolution thus generated when measured from a fixed plane perpendicular to the axis out to a variable plane perpendicular to the axis, show that $D_xV = \pi y^2$.

10. More generally if $A(x)$ denote the area of the section cut from a solid by a plane perpendicular to the x-axis, show that $D_xV = A(x)$.

11. If $A(\phi)$ denote the sectorial area of a plane curve $r = f(\phi)$ and be measured from a fixed radius to a variable radius, show that $D_\phi A = \frac{1}{2}r^2$.

12. If ρ, h, p are the density, height, pressure in a vertical column of air, show that $dp/dh = -\rho$. If $\rho = kp$, show $p = Ce^{-kh}$.

13. Draw a graph to illustrate an apparent exception to the Theorem of the Mean analogous to the apparent exception to Rolle's Theorem, and discuss.

14. Show that the analytic statement of the Theorem of the Mean for $f(x)$ is that a value $x = \xi$ intermediate to a and b may be found such that

$$f(b) - f(a) = f'(\xi)(b - a). \qquad a < \xi < b.$$

15. Show that the semiaxis of an ellipse is a mean proportional between the x-intercept of the tangent and the abscissa of the point of contact.

16. Find the values of the length of the tangent (α) from the point of tangency to the x-axis, (β) to the y-axis, (γ) the total length intercepted between the axes. Consider the same problems for the normal (figure on page 8).

17. Find the angle of intersection of (α) $y^2 = 2mx$ and $x^2 + y^2 = a^2$. (β) $x^2 = 4ay$ and $y = \dfrac{8a^3}{x^2 + 4a^2}$, ($\gamma$) $\dfrac{x^2}{a^2 - \lambda^2} + \dfrac{y^2}{b^2 - \lambda^2} = 1$ for $0 < \lambda < b$ and $b < \lambda < a$.

18. A constant length is laid off along the normal to a parabola. Find the locus.

19. The length of the tangent to $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$ intercepted by the axes is constant.

20. The triangle formed by the asymptotes and any tangent to a hyperbola has constant area.

21. Find the length PT of the tangent to $x = \sqrt{c^2 - y^2} + c\,\mathrm{sech}^{-1}(y/c)$.

22. Find the greatest right cylinder inscribed in a given right cone.

23. Find the cylinder of greatest lateral surface inscribed in a sphere.

24. From a given circular sheet of metal cut out a sector that will form a cone (without base) of maximum volume.

25. Join two points A, B in the same side of a line to a point P of the line in such a way that the distance $PA + PB$ shall be least.

26. Obtain the formula for the distance from a point to a line as the minimum distance.

27. *Test for maximum or minimum.* (α) If $f(x)$ vanishes at the ends of an interval and is positive within the interval and if $f'(x) = 0$ has only one root in the interval, that root indicates a maximum. Prove this by Rolle's Theorem. Apply it in Exs. 22–24. (β) If $f(x)$ becomes indefinitely great at the ends of an interval and $f'(x) = 0$ has only one root in the interval, that root indicates a minimum.

Prove by Rolle's Theorem, and apply in Exs. 25–26. These rules or various modifications of them generally suffice in practical problems to distinguish between maxima and minima without examining either the changes in sign of the first derivative or the sign of the second derivative; for generally there is only one root of $f'(x) = 0$ in the region considered.

28. Show that $x^{-1}\sin x$ from $x = 0$ to $x = \frac{1}{2}\pi$ steadily decreases from 1 to $2/\pi$.

29. If $0 < x < 1$, show (α) $0 < x - \log(1+x) < \dfrac{1}{2}x^2$, (β) $\dfrac{\frac{1}{2}x^2}{1+x} < x - \log(1+x)$.

30. If $0 > x > -1$, show that $\dfrac{1}{2}x^2 < x - \log(1+x) < \dfrac{\frac{1}{2}x^2}{1+x}$.

8. Derivatives of higher order. The derivative of the derivative (regarded as itself a function of x) is the second derivative, and so on to the nth derivative. Customary notations are:

$$f''(x) = \frac{d^2f(x)}{dx^2} = \frac{d^2y}{dx^2} = D_x^2 f = D_x^2 y = y'' = D^2 f = D^2 y,$$

$$f'''(x),\ f^{\text{iv}}(x),\ \cdots,\ f^{(n)}(x);\quad \frac{d^3y}{dx^3},\ \frac{d^4y}{dx^4},\ \cdots,\ \frac{d^ny}{dx^n},\ \cdots.$$

The nth derivative of the sum or difference is the sum or difference of the nth derivatives. For the nth derivative of the product there is a special formula known as *Leibniz's Theorem.* It is

$$D^n(uv) = D^n u \cdot v + nD^{n-1}u\,Dv + \frac{n(n-1)}{2!}D^{n-2}u\,D^2v + \cdots + uD^n v. \quad (44)$$

This result may be written in symbolic form as

$$\text{Leibniz's Theorem}\qquad D^n(uv) = (Du + Dv)^n, \qquad (44')$$

where it is to be understood that in expanding $(Du + Dv)^n$ the term $(Du)^k$ is to be replaced by $D^k u$ and $(Du)^0$ by $D^0 u = u$. In other words the powers refer to repeated differentiations.

A proof of (44) by induction will be found in § 27. The following proof is interesting on account of its ingenuity. Note first that from

$$D(uv) = uDv + vDu, \quad D^2(uv) = D(uDv) + D(vDu),$$

and so on, it appears that $D^2(uv)$ consists of a sum of terms, in each of which there are two differentiations, with numerical coefficients independent of u and v. In like manner it is clear that

$$D^n(uv) = C_0 D^n u \cdot v + C_1 D^{n-1}u\,Dv + \cdots + C_{n-1}Du\,D^{n-1}v + C_n u D^n v$$

is a sum of terms, in each of which there are n differentiations, with coefficients C independent of u and v. To determine the C's any suitable functions u and v, say,

$$u = e^x, \quad v = e^{ax}, \quad uv = e^{(1+a)x}, \quad D^k e^{ax} = a^k e^{ax},$$

may be substituted. If the substitution be made and $e^{(1+a)x}$ be canceled,

$$e^{-(1+a)x}D^n(uv) = (1+a)^n = C_0 + C_1 a + \cdots + C_{n-1}a^{n-1} + C_n a^n,$$

and hence the C's are the coefficients in the binomial expansion of $(1+a)^n$.

Formula (4) for the derivative of a function function may be extended to higher derivatives by repeated applic 1. More generally *any desired change of variable may be made by the* *peated use of* (4) *and* (5). For if x and y be expressed in terms of own functions of new variables u and v, it is always possible to obta the derivatives $D_x y$, $D_x^2 y$, $\cdots$ in terms of $D_u v$, $D_u^2 v$, $\cdots$, and thus an xpression $F(x, y, y', y'', \cdots)$ may be changed into an equivalent ression $\Phi(u, v, v', v'', \cdots)$ in the new variables. In each case that a es the transformations should be carried out by repeated application (4) and (5) rather than by substitution in any general formulas.

The following typical cases are illustrative of the method of change of variab Suppose only the dependent variable y is to be changed to z defined as $y = f(z)$. Th

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}\left(\frac{dz}{dx}\frac{dy}{dz}\right) = \frac{d^2z}{dx^2}\frac{dy}{dz} + \frac{dz}{dx}\left(\frac{d}{dx}\frac{dy}{dz}\right)$$

$$= \frac{d^2z}{dx^2}\frac{dy}{dz} + \frac{dz}{dx}\left(\frac{d}{dz}\frac{dy}{dz}\frac{dz}{dx}\right) = \frac{d^2z}{dx^2}\frac{dy}{dz} + \left(\frac{dz}{dx}\right)^2\frac{d^2y}{dz^2}.$$

As the derivatives of $y = f(z)$ are known, the derivative d^2y/dx^2 has been expr in terms of z and derivatives of z with respect to x. The third derivative wou found by repeating the process. If the problem were to change the indepe variable x to z, defined by $x = f(z)$,

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx} = \frac{dy}{dz}\left(\frac{dx}{dz}\right)^{-1}, \qquad \frac{d^2y}{dx^2} = \frac{d}{dx}\left[\frac{dy}{dz}\left(\frac{dx}{dz}\right)^{-1}\right].$$

$$\frac{d^2y}{dx^2} = \frac{d^2y}{dz^2}\frac{dz}{dx}\left(\frac{dx}{dz}\right)^{-1} - \frac{dy}{dz}\left(\frac{dx}{dz}\right)^{-2}\frac{dz}{dx}\frac{d^2x}{dz^2} = \left[\frac{d^2y}{dz^2}\frac{dx}{dz} - \frac{d^2x}{dz^2}\frac{dy}{dz}\right] \div \Big($$

The change is thus made as far as derivatives of the second order are concer the change of both dependent and independent variables was to be made, th would be similar. Particularly useful changes are to find the derivatives of when y and x are expressed parametrically as functions of t, or when both pressed in terms of new variables r, ϕ as $x = r\cos\phi$, $y = r\sin\phi$. For these see the exercises.

9. The *concavity of a curve* $y = f(x)$ is given by the table:

if $f''(x_0) > 0$,	the curve is concave up at $x = x_0$,
if $f''(x_0) < 0$,	the curve is concave down at $x = x_0$,
if $f''(x_0) = 0$,	an inflection point at $x = x_0$. (?)

Hence the *criterion for distinguishing between maxima and minima*

if $f'(x_0) = 0$ and $f''(x_0) > 0$,	a minimum at $x = x_0$.
if $f'(x_0) = 0$ and $f''(x_0) < 0$,	a maximum at $x = x_0$.
if $f'(x_0) = 0$ and $f''(x_0) = 0$,	neither max. nor min. (?

The question points are necessary in the third line because the statements are not always true unless $f'''(x_0) \neq 0$ (see Ex. 7 under § 39).

It may be recalled that the reason that the curve is concave up in case $f''(x_0) > 0$ is because the derivative $f'(x)$ is then an increasing function in the neighborhood of $x = x_0$; whereas if $f''(x_0) < 0$, the derivative $f'(x)$ is a decreasing function and the curve is convex up. It should be noted that concave up is not the same as concave toward the x-axis, except when the curve is below the axis. With regard to the use of the second derivative as a criterion for distinguishing between maxima and minima, it should be stated that in practical examples the criterion is of relatively small value. It is usually shorter to discuss the change of sign of $f'(x)$ directly, — and indeed in most cases either a rough graph of $f(x)$ or the physical conditions of the problem which calls for the determination of a maximum or minimum will immediately serve to distinguish between them (see Ex. 27 above).

The second derivative is fundamental in dynamics. By definition the *average velocity* $\bar{v}$ of a particle is the ratio of the space traversed to the time consumed, $\bar{v} = s/t$. The *actual velocity* v at any time is the limit of this ratio when the interval of time is diminished and approaches zero as its limit. Thus

$$\bar{v} = \frac{\Delta s}{\Delta t} \quad \text{and} \quad v = \lim_{\Delta t \doteq 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt}. \tag{45}$$

In like manner if a particle describes a straight line, say the x-axis, the *average acceleration* $\bar{f}$ is the ratio of the increment of velocity to the increment of time, and the *actual acceleration* f at any time is the limit of this ratio as $\Delta t \doteq 0$. Thus

$$\bar{f} = \frac{\Delta v}{\Delta t} \quad \text{and} \quad f = \lim_{\Delta t \doteq 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} = \frac{d^2x}{dt^2}. \tag{46}$$

By *Newton's Second Law of Motion, the force acting on the particle is equal to the rate of change of momentum* with the time, momentum being defined as the product of the mass and velocity. Thus

$$F = \frac{d(mv)}{dt} = m\frac{dv}{dt} = mf = m\frac{d^2x}{dt^2}, \tag{47}$$

where it has been assumed in differentiating that the mass is constant, as is usually the case. Hence (47) appears as the fundamental equation for rectilinear motion (see also §§ 79, 84). It may be noted that

$$F = mv\frac{dv}{dx} = \frac{d}{dx}\left(\frac{1}{2}mv^2\right) = \frac{dT}{dx}, \tag{47'}$$

where $T = \frac{1}{2}mv^2$ denotes by definition the *kinetic energy* of the particle. For comments see Ex. 6 following.

EXERCISES

1. State and prove the extension of Leibniz's Theorem to products of three or more factors. Write out the square and cube of a trinomial.

2. Write, by Leibniz's Theorem, the second and third derivatives:

(α) $e^x \sin x$, (β) $\cosh x \cos x$, (γ) $x^2 e^x \log x$.

3. Write the nth derivatives of the following functions, of which the last three should first be simplified by division or separation into partial fractions.

(α) $\sqrt{x+1}$, (β) $\log(ax+b)$, (γ) $(x^2+1)(x+1)^{-3}$,
(δ) $\cos ax$, (ϵ) $e^x \sin x$, (ζ) $(1-x)/(1+x)$,
(η) $\dfrac{1}{x^2-1}$, (θ) $\dfrac{x^3+x+1}{x-1}$, (ι) $\left(\dfrac{ax+1}{ax-1}\right)^2$.

4. If y and x are each functions of t, show that

$$\frac{d^2y}{dx^2} = \frac{\dfrac{dx}{dt}\dfrac{d^2y}{dt^2} - \dfrac{dy}{dt}\dfrac{d^2x}{dt^2}}{\left(\dfrac{dx}{dt}\right)^3} = \frac{x'y'' - y'x''}{x'^3},$$

$$\frac{d^3y}{dx^3} = \frac{x'(x'y''' - y'x''') - 3x''(x'y'' - y'x'')}{x'^5}.$$

5. Find the inflection points of the curve $x = 4\phi - 2\sin\phi$, $y = 4 - 2\cos\phi$.

6. Prove (47′). Hence infer that the force which is the time-derivative of the momentum mv by (47) is also the space-derivative of the kinetic energy.

7. If A denote the area under a curve, as in (43), find $dA/d\theta$ for the curves

(α) $y = a(1-\cos\theta)$, $x = a(\theta - \sin\theta)$, ($\beta$) $x = a\cos\theta$, $y = b\sin\theta$.

8. Make the indicated change of variable in the following equations:

(α) $\dfrac{d^2y}{dx^2} + \dfrac{2x}{1+x^2}\dfrac{dy}{dx} + \dfrac{y}{(1+x^2)^2} = 0$, $x = \tan z$. *Ans.* $\dfrac{d^2y}{dz^2} + y = 0$.

(β) $(1-x^2)\left[\dfrac{d^2y}{dx^2} - \dfrac{1}{y}\left(\dfrac{dy}{dx}\right)^2\right] - x\dfrac{dy}{dx} + y = 0$, $y = e^v$, $x = \sin u$.

Ans. $\dfrac{d^2v}{du^2} + 1 = 0$.

9. *Transformation to polar coördinates.* Suppose that $x = r\cos\phi$, $y = r\sin\phi$. Then

$$\frac{dx}{d\phi} = \frac{dr}{d\phi}\cos\phi - r\sin\phi, \qquad \frac{dy}{d\phi} = \frac{dr}{d\phi}\sin\phi + r\cos\phi,$$

and so on for higher derivatives. Find $\dfrac{dy}{dx}$ and $\dfrac{d^2y}{dx^2} = \dfrac{r^2 + 2(D_\phi r)^2 - rD_\phi^2 r}{(\cos\phi\, D_\phi r - r\sin\phi)^3}$.

10. Generalize formula (5) for the differentiation of an inverse function. Find d^2x/dy^2 and d^3x/dy^3. Note that these may also be found from Ex. 4.

11. A point describes a circle with constant speed. Find the velocity and acceleration of the projection of the point on any fixed diameter.

12. Prove $\dfrac{d^2y}{dx^2} = 2uv^3 + 4v^4\left(\dfrac{dv}{du}\right)^{-1} - v^5\dfrac{d^2v}{du^2}\left(\dfrac{dv}{du}\right)^{-3}$ if $x = \dfrac{1}{v}$, $y = uv$.

10. The indefinite integral. To integrate a function $f(x)$ is to *find a function $F(x)$ the derivative of which is $f(x)$*. The integral $F(x)$ is not uniquely determined by the integrand $f(x)$; for any two functions which differ merely by an additive constant have the same derivative. In giving formulas for integration the constant may be omitted and understood; but in applications of integration to actual problems it should always be inserted and must usually be determined to fit the requirements of special conditions imposed upon the problem and known as the *initial conditions.*

It must not be thought that the constant of integration always appears added to the function $F(x)$. It may be combined with $F(x)$ so as to be somewhat disguised. Thus

$$\log x, \quad \log x + C, \quad \log Cx, \quad \log(x/C)$$

are all integrals of $1/x$, and all except the first have the constant of integration C, although only in the second does it appear as formally additive. To illustrate the determination of the constant by initial conditions, consider the problem of finding the area under the curve $y = \cos x$. By (43)

$$D_x A = y = \cos x \quad \text{and hence} \quad A = \sin x + C.$$

If the area is to be measured from the ordinate $x = 0$, then $A = 0$ when $x = 0$, and by direct substitution it is seen that $C = 0$. Hence $A = \sin x$. But if the area be measured from $x = -\frac{1}{2}\pi$, then $A = 0$ when $x = -\frac{1}{2}\pi$ and $C = 1$. Hence $A = 1 + \sin x$. In fact the area under a curve is not definite until the ordinate from which it is measured is specified, and the constant is needed to allow the integral to fit this initial condition.

11. The fundamental formulas of integration are as follows:

$$\int \frac{1}{x} = \log x, \qquad \int x^n = \frac{1}{n+1} x^n \text{ if } n \neq -1, \tag{48}$$

$$\int e^x = e^x, \qquad \int a^x = a^x/\log a, \tag{49}$$

$$\int \sin x = -\cos x, \qquad \int \cos x = \sin x, \tag{50}$$

$$\int \tan x = -\log\cos x, \qquad \int \cot x = \log\sin x, \tag{51}$$

$$\int \sec^2 x = \tan x, \qquad \int \csc^2 x = -\cot x, \tag{52}$$

$$\int \tan x \sec x = \sec x, \qquad \int \cot x \csc x = -\csc x, \tag{53}$$

with formulas similar to (50)–(53) for the hyperbolic functions. Also

$$\int \frac{1}{1+x^2} = \tan^{-1} x \text{ or } -\cot^{-1} x, \quad \int \frac{1}{1-x^2} = \tanh^{-1} x \text{ or } \coth^{-1} x, \tag{54}$$

$$\int\frac{1}{\sqrt{1-x^2}}=\sin^{-1}x \text{ or } -\cos^{-1}x, \quad \int\frac{\pm 1}{\sqrt{1+x^2}}=\pm\sinh^{-1}x, \tag{55}$$

$$\int\frac{1}{x\sqrt{x^2-1}}=\sec^{-1}x \text{ or } -\csc^{-1}x, \quad \int\frac{\pm 1}{x\sqrt{1-x^2}}=\mp\operatorname{sech}^{-1}x, \tag{56}$$

$$\int\frac{\pm 1}{\sqrt{x^2-1}}=\pm\cosh^{-1}x, \qquad \int\frac{\pm 1}{x\sqrt{1+x^2}}=\mp\operatorname{csch}^{-1}x, \tag{57}$$

$$\int\frac{1}{\sqrt{2x-x^2}}=\operatorname{vers}^{-1}x, \qquad \int\sec x=\operatorname{gd}^{-1}x=\log\tan\left(\frac{\pi}{4}+\frac{x}{2}\right). \tag{58}$$

For the integrals expressed in terms of the inverse hyperbolic functions, the logarithmic equivalents are sometimes preferable. This is not the case, however, in the many instances in which the problem calls for immediate solution with regard to x. Thus if $y=\int(1+x^2)^{-\frac{1}{2}}=\sinh^{-1}x+C$, then $x=\sinh(y-C)$, and the solution is effected and may be translated into exponentials. This is not so easily accomplished from the form $y=\log\left(x+\sqrt{1+x^2}\right)+C$. For this reason and because the inverse hyperbolic functions are briefer and offer striking analogies with the inverse trigonometric functions, it has been thought better to use them in the text and allow the reader to make the necessary substitutions from the table (30)–(35) in case the logarithmic form is desired.

12. In addition to these special integrals, which are consequences of the corresponding formulas for differentiation, there are the general rules of integration which arise from (4) and (6).

$$\int\frac{dz}{dy}\frac{dy}{dx}=\int\frac{dz}{dx}=z, \tag{59}$$

$$\int(u+v-w)=\int u+\int v-\int w, \tag{60}$$

$$uv=\int uv'+\int u'v. \tag{61}$$

Of these rules the second needs no comment and the third will be treated later. Especial attention should be given to the first. For instance suppose it were required to integrate $2\log x/x$. This does not fall under any of the given types; but

$$\frac{2}{x}\log x=\frac{d(\log x)^2}{d\log x}\frac{d\log x}{dx}=\frac{dz}{dy}\frac{dy}{dx}.$$

Here $(\log x)^2$ takes the place of z and $\log x$ takes the place of y. The integral is therefore $(\log x)^2$ as may be verified by differentiation. In general, it may be possible to see that a given integrand is separable into two factors, of which one is integrable when considered as a function of some function of x, while the other is the derivative of that function. Then (59) applies. Other examples are:

$$\int e^{\sin x}\cos x, \qquad \int\tan^{-1}x/(1+x^2), \qquad \int x^2\sin(x^3).$$

In the first, $z = e^y$ is integrable and as $y = \sin x$, $y' = \cos x$; in the second, $z = y$ is integrable and as $y = \tan^{-1}x$, $y' = (1 + x^2)^{-1}$; in the third $z = \sin y$ is integrable and as $y = x^3$, $y' = 3x^2$. The results are

$$e^{\sin x}, \qquad \tfrac{1}{2}(\tan^{-1}x)^2, \qquad -\tfrac{1}{3}\cos(x^3).$$

This method of integration at sight covers such a large percentage of the cases that arise in geometry and physics that it must be thoroughly mastered.*

EXERCISES

1. Verify the fundamental integrals (48)–(58) and give the hyperbolic analogues of (50)–(53).

2. Tabulate the integrals here expressed in terms of inverse hyperbolic functions by means of the corresponding logarithmic equivalents.

3. Write the integrals of the following integrands at sight:

(α) $\sin ax$, (β) $\cot(ax + b)$, (γ) $\tanh 3x$,

(δ) $\dfrac{1}{a^2 + x^2}$, (ϵ) $-\dfrac{1}{\sqrt{x^2 - a^2}}$, (ζ) $\dfrac{1}{\sqrt{2ax - x^2}}$,

(η) $\dfrac{1}{x\log x}$, (θ) $\dfrac{e^{\frac{1}{x}}}{x^2}$, (ι) $\dfrac{x}{x^2 + a^2}$,

(κ) $x^3\sqrt{ax^2 + b}$, (λ) $\tan x \sec^2 x$, (μ) $\cot x \log \sin x$,

(ν) $\dfrac{(x^{-1} - 1)^5}{x^2}$, (o) $\dfrac{\tanh^{-1}x}{1 - x^2}$, (π) $\dfrac{2 + \log x}{x}$,

(ρ) $a^{1 + \sin x}\cos x$, (σ) $-\dfrac{\sin x}{\sqrt{\cos x}}$, (τ) $\dfrac{1}{\sqrt{1 - x^2}\sin^{-1}x}$.

4. Integrate after making appropriate changes such as $\sin^2 x = \frac{1}{2} - \frac{1}{2}\cos 2x$ or $\sec^2 x = 1 + \tan^2 x$, division of denominator into numerator, resolution of the product of trigonometric functions into a sum, completing the square, and so on.

(α) $\cos^2 2x$, (β) $\sin^4 x$. (γ) $\tan^4 x$.

(δ) $\dfrac{1}{x^2 + 3x + 25}$, (ϵ) $\dfrac{2x + 1}{x + 2}$, (ζ) $\dfrac{1 - \sin x}{\operatorname{vers} x}$,

(η) $\dfrac{x + 3}{4x^2 - 5x + 1}$, (θ) $\dfrac{e^{2x} + e^x}{e^{2x} + 1}$, (ι) $\dfrac{1}{\sqrt{2ax + x^2}}$,

(κ) $\sin 5x \cos 2x + 1$. (λ) $\sinh mx \sinh nx$, (μ) $\cos x \cos 2x \cos 3x$,

(ν) $\sec^5 x \tan x - \sqrt{2x}$. (o) $\dfrac{cx + d}{x^2 + ax + b}$, (π) $-\dfrac{x^{m-1}}{(ax^m + b)^p}$.

* The use of differentials (§ 35) is perhaps more familiar than the use of derivatives.

$$z(x) = \int \frac{dz}{dx}dx = \int \frac{dz}{dy}\frac{dy}{dx}dx = \int \frac{dz}{dy}dy = z[y(x)].$$

Then $$\int \frac{2}{x}\log x\,dx = \int 2\log x\,d\log x = (\log x)^2.$$

The use of this notation is left optional with the reader; it has some advantages and some disadvantages. The essential thing is to keep clearly in mind the fact that the problem is to be inspected with a view to detecting the function which will differentiate into the given integrand.

5. How are the following types integrated?

(α) $\sin^m x \cos^n x$, m or n odd, or m and n even,
(β) $\tan^n x$ or $\cot^n x$ when n is an integer,
(γ) $\sec^n x$ or $\csc^n x$ when n is even,
(δ) $\tan^m x \sec^n x$ or $\cot^m x \csc^n x$, n even.

6. Explain the alternative forms in (54)–(56) with all detail possible.

7. Find (α) the area under the parabola $y^2 = 4px$ from $x = 0$ to $x = a$; also (β) the corresponding volume of revolution. Find (γ) the total volume of an ellipsoid of revolution (see Ex. 9, p. 10).

8. Show that the area under $y = \sin mx \sin nx$ or $y = \cos mx \cos nx$ from $x = 0$ to $x = \pi$ is zero if m and n are unequal integers but $\frac{1}{2}\pi$ if they are equal.

9. Find the sectorial area of $r = a \tan \phi$ between the radii $\phi = 0$ and $\phi = \frac{1}{4}\pi$.

10. Find the area of the (α) lemniscate $r^2 = a^2 \cos 2\phi$ and (β) cardioid $r = 1 - \cos\phi$.

11. By Ex. 10, p. 10, find the volumes of these solids. Be careful to choose the parallel planes so that $A(x)$ may be found easily.

(α) The part cut off from a right circular cylinder by a plane through a diameter of one base and tangent to the other. *Ans.* $2/3\pi$ of the whole volume.

(β) How much is cut off from a right circular cylinder by a plane tangent to its lower base and inclined at an angle θ to the plane of the base?

(γ) A circle of radius $b < a$ is revolved, about a line in its plane at a distance a from its center, to generate a ring. The volume of the ring is $2\pi^2ab^2$.

(δ) The axes of two equal cylinders of revolution of radius r intersect at right angles. The volume common to the cylinders is $16r^3/3$.

12. If the cross section of a solid is $A(x) = a_0x^3 + a_1x^2 + a_2x + a_3$, a cubic in x, the volume of the solid between two parallel planes is $\frac{1}{6}h(B + 4M + B')$ where h is the altitude and B and B' are the bases and M is the middle section.

13. Show that $\displaystyle\int \frac{1}{1+x^2} = \tan^{-1}\frac{x+c}{1-cx}$.

13. Aids to integration. The majority of cases of integration which arise in simple applications of calculus may be treated by the method of § 12. Of the remaining cases a large number cannot be integrated at all in terms of the functions which have been treated up to this point. Thus it is impossible to express $\displaystyle\int \frac{1}{\sqrt{(1-x^2)(1-a^2x^2)}}$ in terms of elementary functions. One of the chief reasons for introducing a variety of new functions in higher analysis is to have means for effecting the integrations called for by important applications. The discussion of this matter cannot be taken up here. The problem of integration from an elementary point of view calls for the tabulation of some devices which will accomplish the integration for a

wide variety of integrands integrable in terms of elementary functions. The devices which will be treated are:

Integration by parts, Resolution into partial fractions,

Various substitutions, Reference to tables of integrals.

Integration by parts is an application of (61) when written as

$$\int uv' = uv - \int u'v. \qquad (61')$$

That is, it may happen that the integrand can be written as the product uv' of two factors, where v' is integrable and where $u'v$ is also integrable. Then uv' is integrable. For instance, $\log x$ is not integrated by the fundamental formulas; but

$$\int \log x = \int \log x \cdot 1 = x \log x - \int x/x = x \log x - x.$$

Here $\log x$ is taken as u and 1 as v', so that v is x, u' is $1/x$, and $u'v = 1$ is immediately integrable. This method applies to the inverse trigonometric and hyperbolic functions. Another example is

$$\int x \sin x = - x \cos x + \int \cos x = \sin x - x \cos x.$$

Here if $x = u$ and $\sin x = v'$, both v' and $u'v = - \cos x$ are integrable. If the choice $\sin x = u$ and $x = v'$ had been made, v' would have been integrable but $u'v = \frac{1}{2} x^2 \cos x$ would have been less simple to integrate than the original integrand. Hence in applying integration by parts it is necessary to *look ahead* far enough to see that both v' and $u'v$ are integrable, or at any rate that v' is integrable and the integral of $u'v$ is simpler than the original integral.*

Frequently integration by parts has to be applied several times in succession. Thus

$$\begin{aligned} \int x^2 e^x &= x^2 e^x - \int 2 x e^x && \text{if } u = x^2,\ v' = e^x, \\ &= x^2 e^x - 2\left[x e^x - \int e^x\right] && \text{if } u = x,\ v' = e^x, \\ &= x^2 e^x - 2 x e^x + 2 e^x. \end{aligned}$$

Sometimes it may be applied in such a way as to lead back to the given integral and thus afford an equation from which that integral can be obtained by solution. For example,

$$\begin{aligned} \int e^x \cos x &= e^x \cos x + \int e^x \sin x && \text{if } u = \cos x,\ v' = e^x, \\ &= e^x \cos x + \left[e^x \sin x - \int e^x \cos x\right] && \text{if } u = \sin x,\ v' = e^x, \\ &= e^x (\cos x + \sin x) - \int e^x \cos x. \end{aligned}$$

Hence $$\int e^x \cos x = \tfrac{1}{2} e^x (\cos x + \sin x).$$

* The method of differentials may again be introduced if desired.

14. For the *integration of a rational fraction* $f(x)/F(x)$ where f and F are polynomials in x, the fraction is first resolved into *partial fractions*. This is accomplished as follows. First if f is not of lower degree than F, divide F into f until the remainder is of lower degree than F. The fraction f/F is thus resolved into the sum of a polynomial (the quotient) and a fraction (the remainder divided by F) of which the numerator is of lower degree than the denominator. As the polynomial is integrable, it is merely necessary to consider fractions f/F where f is of lower degree than F. Next it is a fundamental theorem of algebra that a polynomial F may be resolved into linear and quadratic factors

$$F(x) = k(x-a)^{\alpha}(x-b)^{\beta}(x-c)^{\gamma}\cdots(x^2+mx+n)^{\mu}(x^2+px+q)^{\nu}\cdots,$$

where $a, b, c, \cdots$ are the real roots of the equation $F(x) = 0$ and are of the respective multiplicities $\alpha, \beta, \gamma, \cdots$, and where the quadratic factors when set equal to zero give the pairs of conjugate imaginary roots of $F = 0$, the multiplicities of the imaginary roots being $\mu, \nu, \cdots$. It is then a further theorem of algebra that the fraction f/F may be written as

$$\frac{f(x)}{F(x)} = \frac{A_1}{x-a} + \frac{A_2}{(x-a)^2} + \cdots + \frac{A_\alpha}{(x-a)^\alpha} + \frac{B_1}{x-b} + \cdots + \frac{B_\beta}{(x-b)^\beta} + \cdots$$
$$+ \frac{M_1 x + N_1}{x^2+mx+n} + \frac{M_2 x + N_2}{(x^2+mx+n)^2} + \cdots + \frac{M_\mu x + N_\mu}{(x^2+mx+n)^\mu} + \cdots,$$

where there is for each irreducible factor of F a term corresponding to the highest power to which that factor occurs in F and also a term corresponding to every lesser power. The coefficients $A, B, \cdots, M, N, \cdots$ may be obtained by clearing of fractions and equating coefficients of like powers of x, and solving the equations; or they may be obtained by clearing of fractions, substituting for x as many different values as the degree of F, and solving the resulting equations.

When f/F has thus been resolved into partial fractions, the problem has been reduced to the integration of each fraction, and this does not present serious difficulty. The following two examples will illustrate the method of resolution into partial fractions and of integration. Let it be required to integrate

$$\int \frac{x^2+1}{x(x-1)(x-2)(x^2+x+1)} \quad \text{and} \quad \int \frac{2x^3+6}{(x-1)^2(x-3)^3}.$$

The first fraction is expansible into partial fractions in the form

$$\frac{x^2+1}{x(x-1)(x-2)(x^2+x+1)} = \frac{A}{x} + \frac{B}{x-1} + \frac{C}{x-2} + \frac{Dx+E}{x^2+x+1}.$$

Hence $\quad x^2+1 = A(x-1)(x-2)(x^2+x+1) + Bx(x-2)(x^2+x+1)$
$$+ Cx(x-1)(x^2+x+1) + (Dx+E)x(x-1)(x-2).$$

Rather than multiply out and equate coefficients, let $0, 1, 2, -1, -2$ be substituted. Then

$$1 = 2A, \quad 2 = -3B, \quad 5 = 14C, \quad D - E = 1/21, \quad E - 2D = 1/7.$$

$$\int \frac{x^2+1}{x(x-1)(x-2)(x^2+x+1)} = \int \frac{1}{2x} - \int \frac{2}{3(x-1)} + \int \frac{5}{14(x-2)} - \int \frac{4x+5}{21(x^2+x+1)}$$
$$= \frac{1}{2}\log x - \frac{2}{3}\log(x-1) + \frac{5}{14}\log(x-2) - \frac{2}{21}\log(x^2+x+1) - \frac{2}{7\sqrt{3}}\tan^{-1}\frac{2x+1}{\sqrt{3}}.$$

In the second case the form to be assumed for the expansion is

$$\frac{2x^3+6}{(x-1)^2(x-3)^3} = \frac{A}{x-1} + \frac{B}{(x-1)^2} + \frac{C}{(x-3)} + \frac{D}{(x-3)^2} + \frac{E}{(x-3)^3}.$$

$$\begin{aligned} 2x^3+6 = A(x-1)(x-3)^3 + B(x-3)^3 + C(x-1)^2(x-3)^2 \\ + D(x-1)^2(x-3) + E(x-1)^2. \end{aligned}$$

The substitution of 1, 3, 0, 2, 4 gives the equations

$$8 = -8B, \qquad 60 = 4E, \qquad 9A + 3C - D + 12 = 0,$$
$$A - C + D + 6 = 0, \qquad A + 3C + 3D = 0.$$

The solutions are $-9/4$, -1, $+9/4$, $-3/2$, 15, and the integral becomes

$$\begin{aligned} \int \frac{2x^3+6}{(x-1)^2(x-3)^3} = -\frac{9}{4}\log(x-1) + \frac{1}{x-1} + \frac{9}{4}\log(x-3) \\ + \frac{3}{2(x-3)} - \frac{15}{2(x-3)^2}. \end{aligned}$$

The importance of the fact that the method of partial fractions shows that *any rational fraction may be integrated* and, moreover, that the integral may at most consist of a rational part plus the logarithm of a rational fraction plus the inverse tangent of a rational fraction should not be overlooked. Taken with the method of substitution it establishes very wide categories of integrands which are integrable in terms of elementary functions, and effects their integration even though by a somewhat laborious method.

15. The *method of substitution* depends on the identity

$$\int_x f(x) = \int_y f[\phi(y)]\frac{dx}{dy} \qquad \text{if} \qquad x = \phi(y), \tag{59'}$$

which is allied to (59). To show that the integral on the right with respect to y is the integral of $f(x)$ with respect to x it is merely necessary to show that its derivative with respect to x is $f(x)$. By definition of integration,

$$\frac{d}{dy}\int_y f[\phi(y)]\frac{dx}{dy} = f[\phi(y)]\frac{dx}{dy}$$

and

$$\frac{d}{dx}\int_y f[\phi(y)]\frac{dx}{dy} = f[\phi(y)]\frac{dx}{dy}\cdot\frac{dy}{dx} = f[\phi(y)]$$

by (4). The identity is therefore proved. The method of integration by substitution is in fact seen to be merely such a systematization of the method based on (59) and set forth in § 12 as will make it practicable for more complicated problems. Again, differentials may be used if preferred.

Let R denote a rational function. To effect the integration of

$$\int \sin x\, R(\sin^2 x, \cos x), \qquad \text{let} \quad \cos x = y, \qquad \text{then} \quad \int_y -R(1-y^2, y);$$

$$\int \cos x\, R(\cos^2 x, \sin x), \qquad \text{let} \quad \sin x = y, \qquad \text{then} \quad \int_y R(1-y^2, y);$$

$$\int R\left(\frac{\sin x}{\cos x}\right) = \int R(\tan x), \quad \text{let} \quad \tan x = y, \qquad \text{then} \quad \int_y \frac{R(y)}{1+y^2};$$

$$\int R(\sin x, \cos x), \qquad \text{let} \quad \tan\frac{x}{2} = y, \qquad \text{then} \quad \int_y R\left(\frac{2y}{1+y^2}, \frac{1-y^2}{1+y^2}\right)\frac{2}{1+y^2}.$$

The last substitution renders any rational function of $\sin x$ and $\cos x$ rational in the variable y; it should not be used, however, if the previous ones are applicable — it is almost certain to give a more difficult final rational fraction to integrate.

A large number of geometric problems give integrands which are rational in x and in some one of the radicals $\sqrt{a^2+x^2}$, $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$. These may be converted into trigonometric or hyperbolic integrands by the following substitutions:

$$\int R\left(x, \sqrt{a^2-x^2}\right) \qquad x = a\sin y, \qquad \int_y R(a\sin y,\ a\cos y)\,a\cos y;$$

$$\int R\left(x, \sqrt{a^2+x^2}\right) \quad \begin{cases} x = a\tan y, & \int_y R(a\tan y,\ a\sec y)\,a\sec^2 y \\ x = a\sinh y, & \int_y R(a\sinh y,\ a\cosh y)\,a\cosh y; \end{cases}$$

$$\int R\left(x, \sqrt{x^2-a^2}\right) \quad \begin{cases} x = a\sec y, & \int_y R(a\sec y,\ a\tan y)\,a\sec y\tan y \\ x = a\cosh y, & \int_y R(a\cosh y,\ a\sinh y)\,a\sinh y. \end{cases}$$

It frequently turns out that the integrals on the right are easily obtained by methods already given; otherwise they can be treated by the substitutions above.

In addition to these substitutions there are a large number of others which are applied under specific conditions. Many of them will be found among the exercises. Moreover, it frequently happens that an integrand, which does not come under any of the standard types for which substitutions are indicated, is none the less integrable by some substitution which the form of the integrand will suggest.

Tables of integrals, giving the integrals of a large number of integrands, have been constructed by using various methods of integration. B. O. Peirce's "Short Table of Integrals" may be cited. If the particular integrand which is desired does not occur in the Table, it may be possible to devise some substitution which will reduce it to a tabulated form. In the Table are also given a large number of reduction formulas (for the most part deduced by means of integration by parts) which accomplish the successive simplification of integrands which could perhaps be treated by other methods, but only with an excessive amount of labor. Several of these reduction formulas are cited among the exercises. Although the Table is useful in performing integrations and indeed makes it to a large extent unnecessary to learn the various methods of integration, the exercises immediately below, which are constructed for the purpose of illustrating methods of integration, should be done without the aid of a Table.

EXERCISES

1. Integrate the following by parts:

$(\alpha)\ \int x\cosh x,$ $\qquad(\beta)\ \int \tan^{-1}x,$ $\qquad(\gamma)\ \int x^m\log x,$

$(\delta)\ \int \frac{\sin^{-1}x}{x^2},$ $\qquad(\epsilon)\ \int \frac{xe^x}{(1+x)^2},$ $\qquad(\zeta)\ \int \frac{1}{x(x^2-a^2)^{\frac{3}{2}}}.$

2. If $P(x)$ is a polynomial and $P'(x)$, $P''(x)$, $\cdots$ its derivatives, show

$$(\alpha)\ \int P(x)e^{ax} = \frac{1}{a}e^{ax}\left[P(x) - \frac{1}{a}P'(x) + \frac{1}{a^2}P''(x) - \cdots\right],$$

$$(\beta)\ \int P(x)\cos ax = \frac{1}{a}\sin ax\left[P(x) - \frac{1}{a^2}P''(x) + \frac{1}{a^4}P^{\text{iv}}(x) - \cdots\right]$$
$$+ \frac{1}{a}\cos ax\left[\frac{1}{a}P'(x) - \frac{1}{a^3}P'''(x) + \frac{1}{a^5}P^{\text{v}}(x) - \cdots\right],$$

and (γ) derive a similar result for the integrand $P(x)\sin ax$.

3. By successive integration by parts and subsequent solution, show

$$(\alpha)\ \int e^{ax}\sin bx = \frac{e^{ax}(a\sin bx - b\cos bx)}{a^2+b^2},$$

$$(\beta)\ \int e^{ax}\cos bx = \frac{e^{ax}(b\sin bx + a\cos bx)}{a^2+b^2},$$

$$(\gamma)\ \int xe^{2x}\cos x = \tfrac{1}{25}e^{2x}[5x(\sin x + 2\cos x) - 4\sin x - 3\cos x].$$

4. Prove by integration by parts the reduction formulas

$$(\alpha)\ \int \sin^m x\cos^n x = \frac{\sin^{m+1}x\cos^{n-1}x}{m+n} + \frac{n-1}{m+n}\int \sin^m x\cos^{n-2}x,$$

$$(\beta)\ \int \tan^m x\sec^n x = \frac{\tan^{m-1}x\sec^n x}{m+n-1} - \frac{m-1}{m+n-1}\int \tan^{n-2}x\sec^n x,$$

$$(\gamma)\ \int \frac{1}{(x^2+a^2)^n} = \frac{1}{2(n-1)a^2}\left[\frac{x}{(x^2+a^2)^{n-1}} + (2n-3)\int\frac{1}{(x^2+a^2)^{n-1}}\right],$$

$$(\delta)\ \int \frac{x^m}{(\log x)^n} = -\frac{x^{m+1}}{(n-1)(\log x)^{n-1}} + \frac{m+1}{n-1}\int\frac{x^m}{(\log x)^{n-1}}.$$

5. Integrate by decomposition into partial fractions:

$$(\alpha)\ \int\frac{x^2-3x+3}{(x-1)(x-2)},\qquad (\beta)\ \int\frac{1}{a^4-x^4},\qquad (\gamma)\ \int\frac{1}{1+x^4},$$

$$(\delta)\ \int\frac{x^2}{(x+2)^2(x+1)},\qquad (\epsilon)\ \int\frac{4x^2-3x+1}{2x^5+x^3},\qquad (\zeta)\ \int\frac{1}{x(1+x^2)^2}.$$

6. Integrate by trigonometric or hyperbolic substitution:

$$(\alpha)\ \int\sqrt{a^2-x^2},\qquad (\beta)\ \int\sqrt{x^2-a^2},\qquad (\gamma)\ \int\sqrt{a^2+x^2},$$

$$(\delta)\ \int\frac{1}{(a-x^2)^{\frac{3}{2}}},\qquad (\epsilon)\ \int\frac{\sqrt{x^2-a^2}}{x},\qquad (\zeta)\ \int\frac{(a^{\frac{2}{3}}-x^{\frac{2}{3}})^{\frac{3}{2}}}{x^{\frac{1}{3}}}.$$

7. Find the areas of these curves and their volumes of revolution:

$$(\alpha)\ x^{\frac{2}{3}}+y^{\frac{2}{3}}=a^{\frac{2}{3}},\qquad (\beta)\ a^4y^2=a^2x^4-x^6,\qquad (\gamma)\ \left(\frac{x}{a}\right)^{\frac{2}{3}}+\left(\frac{y}{b}\right)^{\frac{2}{3}}=1.$$

8. Integrate by converting to a rational algebraic fraction:

$$(\alpha)\ \int\frac{\sin 3x}{a^2\cos^2x+b^2\sin^2x},\qquad (\beta)\ \int\frac{\cos 3x}{a^2\cos^2x+b^2\sin^2x},\qquad (\gamma)\ \int\frac{\sin 2x}{a^2\cos^2x+b^2\sin^2x},$$

$$(\delta)\ \int\frac{1}{a+b\cos x},\qquad (\epsilon)\ \int\frac{1}{a+b\cos x+c\sin x},\qquad (\zeta)\ \int\frac{1-\cos x}{1+\sin x}.$$

9. Show that $\int R(x, \sqrt{a+bx+cx^2})$ may be treated by trigonometric substitution: distinguish between $b^2-4ac \gtreqless 0$.

10. Show that $\int R\left(x, \sqrt[n]{\dfrac{ax+b}{cx+d}}\right)$ is made rational by $y^n = \dfrac{ax+b}{cx+d}$. Hence infer that $\int R(x, \sqrt{(x-\alpha)(x-\beta)})$ is rationalized by $y^2 = \dfrac{x-\beta}{x-\alpha}$. This accomplishes the integration of $R(x, \sqrt{a+bx+cx^2})$ when the roots of $a+bx+cx^2=0$ are real, that is, when $b^2-4ac>0$.

11. Show that $\int R\left[x, \left(\frac{ax+b}{cx+d}\right)^m, \left(\frac{ax+b}{cx+d}\right)^n, \cdots\right]$, where the exponents m, n, $\cdots$ are rational, is rationalized by $y^k = \frac{ax+b}{cx+d}$ if k is so chosen that km, kn, $\cdots$ are integers.

12. Show that $\int (a+by)^p y^q$ may be rationalized if p or q or $p+q$ is an integer. By setting $x^n = y$ show that $\int x^m (a+bx^n)^p$ may be reduced to the above type and hence is integrable when $\frac{m+1}{n}$ or p or $\frac{m+1}{n}+p$ is integral.

13. If the roots of $a+bx+cx^2=0$ are imaginary, $\int R\left(x, \sqrt{a+bx+cx^2}\right)$ may be rationalized by $y = \sqrt{a+bx+cx^2} \mp x\sqrt{c}$.

14. Integrate the following.

$(\alpha)\ \int \frac{x^3}{\sqrt{x-1}}$, $\quad(\beta)\ \int \frac{1+\sqrt[3]{x}}{1+\sqrt[4]{x}}$, $\quad(\gamma)\ \int \frac{x}{\sqrt[3]{1+x}-\sqrt{1+x}}$,

$(\delta)\ \int \frac{e^{2x}}{\sqrt[4]{e^x+1}}$, $\quad(\epsilon)\ \int \frac{x^4}{\sqrt{(1-x^2)^3}}$, $\quad(\zeta)\ \int \frac{1}{(x-d)\sqrt{a+bx+cx^2}}$,

$(\eta)\ \int \frac{1}{x(1+x^2)^{\frac{3}{2}}}$, $\quad(\theta)\ \int \frac{\sqrt{2x^2+x}}{x^2}$, $\quad(\lambda)\ \int \frac{x^8}{\sqrt{1-x^3}} + \frac{\sqrt{1-x^3}}{x}$.

15. In view of Ex. 12 discuss the integrability of:

$(\alpha)\ \int \sin^m x \cos^n x$, let $\sin x = \sqrt{y}$, $\qquad (\beta)\ \int \frac{x^m}{\sqrt{ax-x^2}}$ $\begin{cases} \text{let } x = ay^2, \\ \text{or } \sqrt{ax-x^2} = xy. \end{cases}$

16. Apply the reduction formulas, Table, p. 66, to show that the final integral for

$$\int \frac{x^m}{\sqrt{1-x^2}} \quad \text{is} \quad \int \frac{1}{\sqrt{1-x^2}} \quad \text{or} \quad \int \frac{x}{\sqrt{1-x^2}} \quad \text{or} \quad \int \frac{1}{x\sqrt{1-x^2}}$$

according as m is even or odd and positive or odd and negative.

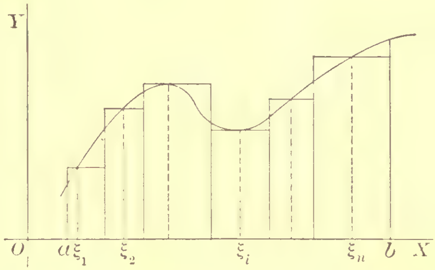
17. Prove sundry of the formulas of Peirce's Table.

18. Show that if $R\left(x, \sqrt{a^2-x^2}\right)$ contains x only to odd powers, the substitution $z = \sqrt{a^2-x^2}$ will rationalize the expression. Use Exs. 1 (ζ) and 6 (ϵ) to compare the labor of this algebraic substitution with that of the trigonometric or hyperbolic.

16. Definite integrals. If an interval from $x=a$ to $x=b$ be divided into n successive intervals $\Delta x_1, \Delta x_2, \cdots, \Delta x_n$ and the value $f(\xi_i)$ of a function $f(x)$ be computed from some point ξ_i in each interval Δx_i and be multiplied by Δx_i, then *the limit of the sum*

$$\lim_{\substack{\Delta x_i \doteq 0 \\ n=\infty}} \left[f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \cdots + f(\xi_n)\Delta x_n\right] = \int_a^b f(x)\,dx. \qquad (62)$$

when each interval becomes infinitely short and their number n becomes infinite, is known as *the definite integral* of $f(x)$ from a to b, and is designated as indicated. If $y=f(x)$ be graphed, the sum will be represented by the area under a broken line, and it is clear that the limit of the sum, that is, the integral, will be represented by the *area under the curve* $y=f(x)$ and between the ordinates $x=a$ and $x=b$. Thus the definite integral, defined arithmetically by (62), may be connected with a geometric concept which can serve to suggest properties of the integral much as the interpretation of the derivative as the slope of the tangent served as a useful geometric representation of the arithmetical definition (2).



For instance, if a, b, c are successive values of x, then

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx \tag{63}$$

is the equivalent of the fact that the area from a to c is equal to the sum of the areas from a to b and b to c. Again, if Δx be considered positive when x moves from a to b, it must be considered negative when x moves from b to a and hence from (62)

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx. \tag{64}$$

Finally, if M be the maximum of $f(x)$ in the interval, the area under the curve will be less than that under the line $y=M$ through the highest point of the curve; and if m be the minimum of $f(x)$, the area under the curve is greater than that under $y=m$. Hence

$$m(b-a) < \int_a^b f(x)\,dx < M(b-a). \tag{65}$$

There is, then, some intermediate value $m<\mu<M$ such that the integral is equal to $\mu(b-a)$; and if the line $y=\mu$ cuts the curve in a point whose abscissa is ξ intermediate between a and b, then

$$\int_a^b f(x)\,dx = \mu(b-a) = (b-a)f(\xi). \tag{65'}$$

This is the fundamental *Theorem of the Mean* for definite integrals.

The definition (62) may be applied directly to the evaluation of the definite integrals of the simplest functions. Consider first $1/x$ and let a, b be positive with a less than b. Let the interval from a to b be divided into n intervals Δx_i which are in geometrical progression in the ratio r so that $x_1 = a$, $x_2 = ar$, $\cdots$, $x_{n+1} = ar^n$

and $\Delta x_1 = a(r-1)$, $\Delta x_2 = ar(r-1)$, $\Delta x_3 = ar^2(r-1)$, $\cdots$, $\Delta x_n = ar^{n-1}(r-1)$;

whence $\quad b - a = \Delta x_1 + \Delta x_2 + \cdots + \Delta x_n = a(r^n - 1) \quad$ and $\quad r^n = b/a$.

Choose the points ξ_i in the intervals Δx_i as the initial points of the intervals. Then

$$\frac{\Delta x_1}{\xi_1} + \frac{\Delta x_2}{\xi_2} + \cdots + \frac{\Delta x_n}{\xi_n} = \frac{a(r-1)}{a} + \frac{ar(r-1)}{ar} + \cdots + \frac{ar^{n-1}(r-1)}{ar^{n-1}} = n(r-1).$$

But $$r = \sqrt[n]{b/a} \quad \text{or} \quad n = \log(b/a) \div \log r.$$

Hence $$\frac{\Delta x_1}{\xi_1} + \frac{\Delta x_2}{\xi_2} + \cdots + \frac{\Delta x_n}{\xi_n} = n(r-1) = \log\frac{b}{a}\cdot\frac{r-1}{\log r} = \log\frac{b}{a}\cdot\frac{h}{\log(1+h)}.$$

Now if n becomes infinite, r approaches 1, and h approaches 0. But the limit of $\log(1+h)/h$ as $h \doteq 0$ is by definition the derivative of $\log(1+x)$ when $x = 0$ and is 1. Hence

$$\int_a^b \frac{dx}{x} = \lim_{n=\infty}\left[\frac{\Delta x_1}{\xi_1} + \frac{\Delta x_2}{\xi_2} + \cdots + \frac{\Delta x_n}{\xi_n}\right] = \log\frac{b}{a} = \log b - \log a.$$

As another illustration let it be required to evaluate the integral of $\cos^2 x$ from 0 to $\frac{1}{2}\pi$. Here let the intervals Δx_i be equal and their number odd. Choose the ξ's as the initial points of their intervals. The sum of which the limit is desired is

$$\sigma = \cos^2 0\cdot\Delta x + \cos^2\Delta x\cdot\Delta x + \cos^2 2\,\Delta x\cdot\Delta x + \cdots$$
$$+ \cos^2(n-2)\,\Delta x\cdot\Delta x + \cos^2(n-1)\,\Delta x\cdot\Delta x.$$

But $\quad n\Delta x = \frac{1}{2}\pi$, and $(n-1)\,\Delta x = \frac{1}{2}\pi - \Delta x$, $(n-2)\,\Delta x = \frac{1}{2}\pi - 2\,\Delta x$, $\cdots$,

and $$\cos(\tfrac{1}{2}\pi - y) = \sin y \quad\text{and}\quad \sin^2 y + \cos^2 y = 1.$$

Hence $$\sigma = \Delta x\left[\cos^2 0 + \cos^2\Delta x + \cos^2 2\,\Delta x + \cdots + \sin^2 2\,\Delta x + \sin^2\Delta x\right]$$
$$= \Delta x\left[1 + \frac{n-1}{2}\right].$$

Hence $$\int_0^{\frac{\pi}{2}}\cos^2 x\,dx = \lim_{\Delta x \doteq 0}\left[\tfrac{1}{2}n\,\Delta x + \tfrac{1}{2}\Delta x\right] = \lim_{\Delta x \doteq 0}\left(\tfrac{1}{4}\pi + \tfrac{1}{2}\Delta x\right) = \tfrac{1}{4}\pi.$$

Indications for finding the integrals of other functions are given in the exercises.

It should be noticed that the variable x which appears in the expression of the definite integral really has nothing to do with the value of the integral but merely serves as a symbol useful in forming the sum in (62). What is of importance is the function f and the limits a, b of the interval over which the integral is taken.

$$\int_a^b f(x)\,dx = \int_a^b f(t)\,dt = \int_a^b f(y)\,dy = \int_a^b f(\times)\,d\times.$$

The variable in the integrand disappears in the integration and leaves the value of the integral as a function of the limits a and b alone.

17. If the lower limit of the integral be fixed, the value

$$\int_a^b f(x)\,dx = \Phi(b)$$

of the integral is a function of the upper limit regarded as variable. To find the derivative $\Phi'(b)$, form the quotient (2),

$$\frac{\Phi(b+\Delta b)-\Phi(b)}{\Delta b} = \frac{\int_a^{b+\Delta b} f(x)\,dx - \int_a^b f(x)\,dx}{\Delta b}.$$

By applying (63) and (65′), this takes the simpler form

$$\frac{\Phi(b+\Delta b)-\Phi(b)}{\Delta b} = \frac{\int_b^{b+\Delta b} f(x)\,dx}{\Delta b} = \frac{1}{\Delta b}\cdot f(\xi)\,\Delta b,$$

where ξ is intermediate between b and $b+\Delta b$. Let $\Delta b \doteq 0$. Then ξ approaches b and $f(\xi)$ approaches $f(b)$. Hence

$$\Phi'(b) = \frac{d}{db}\int_a^b f(x)\,dx = f(b). \tag{66}$$

If preferred, the variable b may be written as x, and

$$\Phi(x) = \int_a^x f(x)\,dx, \qquad \Phi'(x) = \frac{d}{dx}\int_a^x f(x)\,dx = f(x). \tag{66′}$$

This equation will establish the relation between the definite integral and the indefinite integral. For by definition, the indefinite integral $F(x)$ of $f(x)$ is any function such that $F'(x)$ equals $f(x)$. As $\Phi'(x) = f(x)$ it follows that

$$\int_a^x f(x)\,dx = F(x) + C. \tag{67}$$

Hence except for an additive constant, the indefinite integral of f is the definite integral of f from a fixed lower limit to a variable upper limit. As the definite integral vanishes when the upper limit coincides with the lower, the constant C is $-F(a)$ and

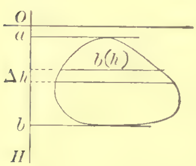
$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{67′}$$

Hence, *the definite integral of $f(x)$ from a to b is the difference between the values of any indefinite integral $F(x)$ taken for the upper and lower limits of the definite integral*; and if the indefinite integral of f is known, the definite integral may be obtained without applying the definition (62) to f.

The great importance of definite integrals to geometry and physics lies in that fact that *many quantities* connected with geometric figures or physical bodies *may be expressed simply for small portions* of the figures or bodies and may then be obtained as the sum of those quantities taken over all the small portions, or rather, as the *limit of that sum when the portions become smaller and smaller.* Thus the area under a curve cannot in the first instance be evaluated; but if only that portion of the curve which lies over a small interval Δx be considered and the rectangle corresponding to the ordinate $f(\xi)$ be drawn, it is clear that the area of the rectangle is $f(\xi)\,\Delta x$, that the area of all the rectangles is the sum $\Sigma f(\xi)\,\Delta x$ taken from a to b, that when the intervals Δx approach zero the limit of their sum is the area under the curve; and hence that area may be written as the definite integral of $f(x)$ from a to b.*

In like manner consider *the mass of a rod* of variable density and suppose the rod to lie along the x-axis so that the density may be taken as a function of x. In any small length Δx of the rod the density is nearly constant and the mass of that part is approximately equal to the product $\rho\Delta x$ of the density $\rho(x)$ at the initial point of that part times the length Δx of the part. In fact it is clear that the mass will be intermediate between the products $m\Delta x$ and $M\Delta x$, where m and M are the minimum and maximum densities in the interval Δx. In other words the mass of the section Δx will be exactly equal to $\rho(\xi)\,\Delta x$ where ξ is some value of x in the interval Δx. The mass of the whole rod is therefore the sum $\Sigma\rho(\xi)\Delta x$ taken from one end of the rod to the other, and if the intervals be allowed to approach zero, the mass may be written as the integral of $\rho(x)$ from one end of the rod to the other.†

Another problem that may be treated by these methods is that of finding the *total pressure* on a vertical area submerged in a liquid, say, in water. Let w be the weight of a column of water of cross section 1 sq. unit and of height 1 unit. (If the unit is a foot, $w = 62.5$ lb.) At a point h units below the surface of the water the pressure is wh and upon a small area near that depth the pressure is approximately whA if A be the area. The pressure on the area A is exactly equal to $w\xi A$ if ξ is some depth intermediate between that of the top and that of the bottom of the area. Now let the finite area be ruled into strips of height Δh. Consider the product $whb(h)\,\Delta h$ where $b(h) = f(h)$ is the breadth of the area at the depth h. This



* The ξ's may evidently be so chosen that the finite sum $\Sigma f(\xi)\Delta x$ is exactly equal to the area under the curve; but still it is necessary to let the intervals approach zero and thus replace the sum by an integral because the values of ξ which make the sum equal to the area are unknown.

† This and similar problems, here treated by using the Theorem of the Mean for integrals, may be treated from the point of view of differentiation as in § 7 or from that of Duhamel's or Osgood's Theorem as in §§ 34, 35. It should be needless to state that in any particular problem some one of the three methods is likely to be somewhat preferable to either of the others. The reason for laying such emphasis upon the Theorem of the Mean here and in the exercises below is that the theorem is in itself very important and needs to be thoroughly mastered.

is approximately the pressure on the strip as it is the pressure at the top of the strip multiplied by the approximate area of the strip. Then $w\xi b(\xi)\Delta h$, where ξ is some value between h and $h + \Delta h$, is the actual pressure on the strip. (It is sufficient to write the pressure as approximately $whb(h)\Delta h$ and not trouble with the ξ.) The total pressure is then $\Sigma w\xi b(\xi)\Delta h$ or better the limit of that sum. Then

$$P = \lim \sum w\xi b(\xi)\,dh = \int_a^b whb(h)\,dh,$$

where a is the depth of the top of the area and b that of the bottom. To evaluate the pressure it is merely necessary to find the breadth b as a function of h and integrate.

EXERCISES

1. If k is a constant, show $\int_a^b kf(x)\,dx = k\int_a^b f(x)\,dx$.

2. Show that $\int_a^b (u \pm v)\,dx = \int_a^b u\,dx \pm \int_a^b v\,dx$.

3. If, from a to b, $\psi(x) < f(x) < \phi(x)$, show $\int_a^b \psi(x)\,dx < \int_a^b f(x)\,dx < \int_a^b \phi(x)\,dx$.

4. Suppose that the minimum and maximum of the quotient $Q(x) = f(x)/\phi(x)$ of two functions in the interval from a to b are m and M, and let $\phi(x)$ be positive so that

$$m < Q(x) = \frac{f(x)}{\phi(x)} < M \quad \text{and} \quad m\phi(x) < f(x) < M\phi(x)$$

are true relations. Show by Exs. 3 and 1 that

$$m < \frac{\int_a^b f(x)\,dx}{\int_a^b \phi(x)\,dx} < M \quad \text{and} \quad \frac{\int_a^b f(x)\,dx}{\int_a^b \phi(x)\,dx} = \mu = Q(\xi) = \frac{f(\xi)}{\phi(\xi)},$$

where ξ is some value of x between a and b.

5. If m and M are the minimum and maximum of $f(x)$ between a and b and if $\phi(x)$ is always positive in the interval, show that

$$m\int_a^b \phi(x)\,dx < \int_a^b f(x)\,\phi(x)\,dx < M\int_a^b \phi(x)\,dx$$

and
$$\int_a^b f(x)\,\phi(x)\,dx = \mu\int_a^b \phi(x)\,dx = f(\xi)\int_a^b \phi(x)\,dx.$$

Note that the integrals of $[M - f(x)]\,\phi(x)$ and $[f(x) - m]\,\phi(x)$ are positive and apply Ex. 2.

6. Evaluate the following by the direct application of (62):

$$(\alpha)\ \int_a^b x\,dx = \frac{b^2 - a^2}{2}, \qquad (\beta)\ \int_a^b e^x dx = e^b - e^a.$$

Take equal intervals and use the rules for arithmetic and geometric progressions.

7. Evaluate $(\alpha)\ \int_a^b x^m dx = \frac{1}{m+1}(b^{m+1} - a^{m+1})$, $\quad(\beta)\ \int_a^b c^x dx = \frac{1}{\log c}(c^b - c^a)$.
In the first the intervals should be taken in geometric progression with $r^n = b/a$.

8. Show directly that (α) $\int_0^\pi \sin^2 x dx = \frac{1}{2}\pi$, (β) $\int_0^\pi \cos^n x dx = 0$, if n is odd.

9. With the aid of the trigonometric formulas

$$\cos x + \cos 2x + \cdots + \cos(n-1)x = \tfrac{1}{2}[\sin nx \cot \tfrac{1}{2}x - 1 - \cos nx],$$
$$\sin x + \sin 2x + \cdots + \sin(n-1)x = \tfrac{1}{2}[(1 - \cos nx)\cot \tfrac{1}{2}x - \sin nx],$$

show (α) $\int_a^b \cos x dx = \sin b - \sin a$, (β) $\int_a^b \sin x dx = \cos a - \cos b$.

10. A function is said to be *even* if $f(-x) = f(x)$ and *odd* if $f(-x) = -f(x)$. Show (α) $\int_{-a}^{+a} f(x)\,dx = 2\int_0^a f(x)\,dx$, f even, (β) $\int_{-a}^{+a} f(x)\,dx = 0$, f odd.

11. Show that if an integral is regarded as a function of the lower limit, the upper limit being fixed, then

$$\Phi'(a) = \frac{d}{da}\int_a^b f(x)\,dx = -f(a), \quad \text{if} \quad \Phi(a) = \int_a^b f(x)\,dx.$$

12. Use the relation between definite and indefinite integrals to compare

$$\int_a^b f(x)\,dx = (b-a)f(\xi) \quad \text{and} \quad F(b) - F(a) = (b-a)F'(\xi),$$

the Theorem of the Mean for derivatives and for definite integrals.

13. From consideration of Exs. 12 and 4 establish *Cauchy's Formula*

$$\frac{\Delta F}{\Delta \Phi} = \frac{F(b) - F(a)}{\Phi(b) - \Phi(a)} = \frac{F'(\xi)}{\Phi'(\xi)}, \qquad a < \xi < b,$$

which states that the quotient of the increments ΔF and $\Delta \Phi$ of two functions, in any interval in which the derivative $\Phi'(x)$ does not vanish, is equal to the quotient of the derivatives of the functions for some interior point of the interval. What would the application of the Theorem of the Mean for derivatives to numerator and denominator of the left-hand fraction give, and wherein does it differ from Cauchy's Formula?

14. Discuss the volume of revolution of $y = f(x)$ as the limit of the sum of thin cylinders and compare the results with those found in Ex. 9, p. 10.

15. Show that the mass of a rod running from a to b along the x-axis is $\frac{1}{2}k(b^2 - a^2)$ if the density varies as the distance from the origin (k is a factor of proportionality).

16. Show (α) that the mass in a rod running from a to b is the same as the area under the curve $y = \rho(x)$ between the ordinates $x = a$ and $x = b$, and explain why this should be seen intuitively to be so. Show (β) that if the density in a plane slab bounded by the x-axis, the curve $y = f(x)$, and the ordinates $x = a$ and $x = b$ is a function $\rho(x)$ of x alone, the mass of the slab is $\int_a^b y\rho(x)\,dx$; also (γ) that the mass of the corresponding volume of revolution is $\int_a^b \pi y^2 \rho(x)\,dx$.

17. An isosceles triangle has the altitude a and the base $2b$. Find (α) the mass on the assumption that the density varies as the distance from the vertex (measured along the altitude). Find (β) the mass of the cone of revolution formed by revolving the triangle about its altitude if the law of density is the same.

18. In a plane, the *moment of inertia I of a particle of mass m with respect to a point* is defined as the product mr^2 of the mass by the square of its distance from the point. Extend this definition from particles to bodies.

(α) Show that the moments of inertia of a rod running from a to b and of a circular slab of radius a are respectively

$$I = \int_a^b x^2\rho(x)\,dx \quad \text{and} \quad I = \int_0^a 2\pi r^3\rho(r)\,dr, \qquad \rho \text{ the density},$$

if the point of reference for the rod is the origin and for the slab is the center.

(β) Show that for a rod of length $2l$ and of uniform density, $I = \frac{1}{3}Ml^2$ with respect to the center and $I = \frac{4}{3}Ml^2$ with respect to the end, M being the total mass of the rod.

(γ) For a uniform circular slab with respect to the center $I = \frac{1}{2}Ma^2$.

(δ) For a uniform rod of length $2l$ with respect to a point at a distance d from its center is $I = M(\frac{1}{3}l^2 + d^2)$. Take the rod along the axis and let the point be (α, β) with $d^2 = \alpha^2 + \beta^2$.

19. A rectangular gate holds in check the water in a reservoir. If the gate is submerged over a vertical distance H and has a breadth B and the top of the gate is a units below the surface of the water, find the pressure on the gate. At what depth in the water is the point where the pressure is the mean pressure over the gate?

20. A dam is in the form of an isosceles trapezoid 100 ft. along the top (which is at the water level) and 60 ft. along the bottom and 30 ft. high. Find the pressure in tons.

21. Find the pressure on a circular gate in a water main if the radius of the circle is r and the depth of the center of the circle below the water level is $d \geqq r$.

22. In space, *moments of inertia* are defined *relative to an axis* and in the formula $I = mr^2$, for a single particle, r is the perpendicular distance from the particle to the axis.

(α) Show that if the density in a solid of revolution generated by $y = f(x)$ varies only with the distance along the axis, the moment of inertia about the axis of revolution is $I = \int_a^b \frac{1}{2}\pi y^4\rho(x)\,dx$. Apply Ex. 18 after dividing the solid into disks.

(β) Find the moment of inertia of a sphere about a diameter in case the density is constant; $I = \frac{2}{5}Ma^2 = \frac{8}{15}\pi\rho a^5$.

(γ) Apply the result to find the moment of inertia of a spherical shell with external and internal radii a and b: $I = \frac{2}{5}M(a^5 - b^5)/(a^3 - b^3)$. Let $b \doteq a$ and thus find $I = \frac{2}{3}Ma^2$ as the moment of inertia of a spherical surface (shell of negligible thickness).

(δ) For a cone of revolution $I = \frac{3}{10}Ma^2$ where a is the radius of the base.

23. If the force of attraction exerted by a mass m upon a point is $kmf(r)$ where r is the distance from the mass to the point, show that the attraction exerted at the origin by a rod of density $\rho(x)$ running from a to b along the x-axis is

$$A = \int_a^b kf(x)\rho(x)\,dx, \quad \text{and that} \quad A = kM/ab, \qquad M = \rho(b - a),$$

is the attraction of a uniform rod if the law is the Law of Nature, that is, $f(r) = 1/r^2$.

24. Suppose that the density ρ in the slab of Ex. 16 were a function $\rho(x, y)$ of both x and y. Show that the mass of a small slice over the interval Δx_i would be of the form

$$\Delta x \int_0^{y=f(\xi)} \rho(x, y)\, dy = \Phi(\xi)\,\Delta x \quad \text{and that} \quad \int_a^b \Phi(x)\,\Delta x = \int_a^b \left[\int_0^{y=f(x)} \rho(x, y)\, dy\right] dx$$

would be the expression for the total mass and would require an integration with respect to y in which x was held constant, a substitution of the limits $f(x)$ and 0 for y, and then an integration with respect to x from a to b.

25. Apply the considerations of Ex. 24 to finding moments of inertia of
(α) a uniform triangle $y = mx$, $y = 0$, $x = a$ with respect to the origin,
(β) a uniform rectangle with respect to the center,
(γ) a uniform ellipse with respect to the center.

26. Compare Exs. 24 and 16 to treat the volume under the surface $z = \rho(x, y)$ and over the area bounded by $y = f(x)$, $y = 0$, $x = a$, $x = b$. Find the volume
(α) under $z = xy$ and over $y^2 = 4px$, $y = 0$, $x = 0$, $x = b$,
(β) under $z = x^2 + y^2$ and over $x^2 + y^2 = a^2$, $y = 0$, $x = 0$, $x = Q$,
(γ) under $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$ and over $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, $y = 0$, $x = 0$, $x = a$.

27. Discuss sectorial area $\frac{1}{2}\int r^2 d\phi$ in polar coördinates as the limit of the sum of small sectors running out from the pole.

28. Show that the moment of inertia of a uniform circular sector of angle α and radius a is $\frac{1}{4}\rho\alpha a^4$. Hence infer $I = \frac{1}{4}\rho\int_{\alpha_0}^{\alpha_1} r^4 d\phi$ in polar coördinates.

29. Find the moment of inertia of a uniform (α) lemniscate $r^2 = a^2\cos^2 2\phi$ and (β) cardioid $r = a(1 - \cos\phi)$ with respect to the pole. Also of (γ) the circle $r = 2a\cos\phi$ and (δ) the rose $r = a\sin 2\phi$ and (ϵ) the rose $r = a\sin 3\phi$.

CHAPTER II

REVIEW OF FUNDAMENTAL THEORY*

18. Numbers and limits. The concept and theory of *real number*, integral, rational, and irrational, will not be set forth in detail here. Some matters, however, which are necessary to the proper understanding of rigorous methods in analysis must be mentioned; and numerous points of view which are adopted in the study of irrational number will be suggested in the text or exercises.

It is taken for granted that by his earlier work the reader has become familiar with the use of real numbers. In particular it is assumed that he is accustomed to represent numbers as a *scale*, that is, by points on a straight line, and that he knows that when a line is given and an origin chosen upon it and a unit of measure and a positive direction have been chosen, then to each point of the line corresponds one and only one real number, and conversely. Owing to this correspondence, that is, owing to the conception of a scale, it is possible to interchange statements about numbers with statements about points and hence to obtain a more vivid and graphic or a more abstract and arithmetic phraseology as may be desired. Thus instead of saying that the numbers $x_1, x_2, \cdots$ are increasing algebraically, one may say that the points (whose coördinates are) $x_1, x_2, \cdots$ are moving in the positive direction or to the right; with a similar correlation of a decreasing suite of numbers with points moving in the negative direction or to the left. It should be remembered, however, that whether a statement is couched in geometric or algebraic terms, it is always a statement concerning numbers when one has in mind the point of view of pure analysis.†

It may be recalled that arithmetic begins with the integers, including 0, and with addition and multiplication. That second, the rational numbers of the form p/q are introduced with the operation of division and the negative rational numbers with the operation of subtraction. Finally, the irrational numbers are introduced by various processes. Thus $\sqrt{2}$ occurs in geometry through the necessity of expressing the length of the diagonal of a square, and $\sqrt{3}$ for the diagonal of a cube. Again, π is needed for the ratio of circumference to diameter in a circle. In algebra any equation of odd degree has at least one real root and hence may be regarded as defining a number. But there is an essential difference between rational and irrational numbers in that any rational number is of the

* The object of this chapter is to set forth systematically, with attention to precision of statement and accuracy of proof, those fundamental definitions and theorems which lie at the basis of calculus and which have been given in the previous chapter from an intuitive rather than a critical point of view.

† Some illustrative graphs will be given; the student should make many others.

form $\pm p/q$ with $q \neq 0$ and can therefore be written down explicitly; whereas the irrational numbers arise by a variety of processes and, although they may be represented to any desired accuracy by a decimal, they cannot all be written down explicitly. It is therefore necessary to have some definite axioms regulating the essential properties of irrational numbers. The particular axiom upon which stress will here be laid is the axiom of continuity, the use of which is essential to the proof of elementary theorems on limits.

19. Axiom of Continuity. *If all the points of a line are divided into two classes such that every point of the first class precedes every point of the second class, there must be a point C such that any point preceding C is in the first class and any point succeeding C is in the second class.* This principle may be stated in terms of numbers, as: *If all real numbers be assorted into two classes such that every number of the first class is algebraically less than every number of the second class, there must be a number N such that any number less than N is in the first class and any number greater than N is in the second.* The number N (or point C) is called the frontier number (or point), or simply the *frontier* of the two classes, and in particular it is the *upper frontier* for the first class and the *lower frontier* for the second.

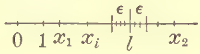
To consider a particular case, let all the negative numbers and zero constitute the first class and all the positive numbers the second, or let the negative numbers alone be the first class and the positive numbers with zero the second. In either case it is clear that the classes satisfy the conditions of the axiom and that zero is the frontier number such that any lesser number is in the first class and any greater in the second. If, however, one were to consider the system of all positive and negative numbers but without zero, it is clear that there would be no number N which would satisfy the conditions demanded by the axiom when the two classes were the negative and positive numbers; for no matter how small a positive number were taken as N, there would be smaller numbers which would also be positive and would not belong to the first class; and similarly in case it were attempted to find a negative N. Thus the axiom insures the presence of zero in the system, and in like manner insures the presence of every other number — a matter which is of importance because there is no way of writing all (irrational) numbers in explicit form.

Further to appreciate the continuity of the number scale, consider the four significations attributable to the phrase "*the interval from a to b.*" They are

$$a \leqq x \leqq b, \qquad a < x \leqq b, \qquad a \leqq x < b, \qquad a < x < b.$$

That is to say, both end points or either or neither may belong to the interval. In the case a is absent, the interval has no first point; and if b is absent, there is no last point. Thus if zero is not counted as a positive number, there is no least positive number; for if any least number were named, half of it would surely be less, and hence the absurdity. The axiom of continuity shows that if all numbers be divided into two classes as required, there must be either a greatest in the first class or a least in the second — the frontier — but not both unless the frontier is counted twice, once in each class.

20. Definition of a Limit. *If x is a variable which takes on successive values $x_1, x_2, \cdots, x_i, x_j, \cdots$, the variable x is said to approach the constant l as a limit if the numerical difference between x and l ultimately becomes, and for all succeeding values of x remains, less than any preassigned number no matter how small.* The numerical difference between x and l is denoted by $|x-l|$ or $|l-x|$ and is called the *absolute value* of the difference. The fact of the approach to a limit may be stated as



$$|x-l| < \epsilon \quad \text{for all } x\text{'s subsequent to some } x$$

or $$x = l + \eta, \quad |\eta| < \epsilon \quad \text{for all } x\text{'s subsequent to some } x,$$

where ϵ is a positive number which may be assigned at pleasure and must be assigned before the attempt be made to find an x such that for all subsequent x's the relation $|x-l| < \epsilon$ holds.

So long as the conditions required in the definition of a limit are satisfied there is no need of bothering about how the variable approaches its limit, whether from one side or alternately from one side and the other, whether discontinuously as in the case of the area of the polygons used for computing the area of a circle or continuously as in the case of a train brought to rest by its brakes. To speak geometrically, a point x which changes its position upon a line approaches the point l as a limit if the point x ultimately comes into and remains in an assigned interval, no matter how small, surrounding l.

A variable is said to *become infinite* if the numerical value of the variable ultimately becomes and remains greater than any preassigned number K, no matter how large.* The notation is $x = \infty$, but had best be read "x becomes infinite," not "x equals infinity."

Theorem 1. If a variable is always increasing, it either becomes infinite or approaches a limit.

That the variable *may* increase indefinitely is apparent. But if it does not become infinite, there must be numbers K which are greater than any value of the variable. Then any number must satisfy one of two conditions: either there are values of the variable which are greater than it or there are no values of the variable greater than it. Moreover all numbers that satisfy the first condition are less than any number which satisfies the second. All numbers are therefore divided into two classes fulfilling the requirements of the axiom of continuity, and there must be a number N such that there are values of the variable greater than any number $N-\epsilon$ which is less than N. Hence if ϵ be assigned, there is a value of the variable which lies in the interval $N-\epsilon < x \leqq N$, and as the variable is always increasing, all subsequent values must lie in this interval. Therefore the variable approaches N as a limit.

* This definition means what it says, and no more. Later, additional or different meanings may be assigned to infinity, but not now. Loose and extraneous concepts in this connection are almost certain to introduce errors and confusion.

EXERCISES

1. If $x_1, x_2, \cdots, x_n, \cdots, x_{n+p}, \cdots$ is a suite approaching a limit, apply the definition of a limit to show that when ϵ is given it must be possible to find a value of n so great that $|x_{n+p} - x_n| < \epsilon$ for all values of p.

2. If $x_1, x_2, \cdots$ is a suite approaching a limit and if $y_1, y_2, \cdots$ is any suite such that $|y_n - x_n|$ approaches zero when n becomes infinite, show that the y's approach a limit which is identical with the limit of the x's.

3. As the definition of a limit is phrased in terms of inequalities and absolute values, note the following rules of operation:

$$(\alpha)\ \text{If}\quad a > 0 \quad\text{and}\quad c > b,\ \text{then}\quad \frac{c}{a} > \frac{b}{a} \quad\text{and}\quad \frac{a}{c} < \frac{a}{b},$$

$$(\beta)\ |a + b + c + \cdots| \leqq |a| + |b| + |c| + \cdots, \qquad (\gamma)\ |abc \cdots| = |a| \cdot |b| \cdot |c| \cdots,$$

where the equality sign in (β) holds only if the numbers $a, b, c, \cdots$ have the same sign. By these relations and the definition of a limit prove the fundamental theorems:

If $\lim x = X$ and $\lim y = Y$, then $\lim (x \pm y) = X \pm Y$ and $\lim xy = XY$.

4. Prove Theorem 1 when restated in the slightly changed form: If a variable x never decreases and never exceeds K, then x approaches a limit N and $N \leqq K$. Illustrate fully. State and prove the corresponding theorem for the case of a variable never increasing.

5. If $x_1, x_2, \cdots$ and $y_1, y_2, \cdots$ are two suites of which the first never decreases and the second never increases, all the y's being greater than any of the x's, and if when ϵ is assigned an n can be found such that $y_n - x_n < \epsilon$, show that the limits of the suites are identical.

6. If $x_1, x_2, \cdots$ and $y_1, y_2, \cdots$ are two suites which never decrease, show by Ex. 4 (not by Ex. 3) that the suites $x_1 + y_1, x_2 + y_2, \cdots$ and $x_1y_1, x_2y_2, \cdots$ approach limits. Note that two infinite decimals are precisely two suites which never decrease as more and more figures are taken. They do not always increase, for some of the figures may be 0.

7. If the word "all" in the hypothesis of the axiom of continuity be assumed to refer only to rational numbers so that the statement becomes: If all rational numbers be divided into two classes $\cdots$, there shall be a number N (not necessarily rational) such that $\cdots$; then the conclusion may be taken as defining a number as the frontier of a sequence of rational numbers. Show that if two numbers X, Y be defined by two such sequences, and if the sum of the numbers be *defined* as the number defined by the sequence of the sums of corresponding terms as in Ex. 6, and if the product of the numbers be *defined* as the number defined by the sequence of the products as in Ex. 6, then the fundamental rules

$$X + Y = Y + X, \qquad XY = YX, \qquad (X + Y)Z = XZ + YZ$$

of arithmetic hold for the numbers X, Y, Z defined by sequences. In this way a complete theory of irrationals may be built up from the properties of rationals combined with the principle of continuity, namely, 1° by defining irrationals as frontiers of sequences of rationals, 2° by defining the operations of addition, multiplication, $\cdots$ as operations upon the rational numbers in the sequences, 3° by showing that the fundamental rules of arithmetic still hold for the irrationals.

8. Apply the principle of continuity to show that there is a positive number x such that $x^2 = 2$. To do this it should be shown that the rationals are divisible into two classes, those whose square is less than 2 and those whose square is not less than 2; and that these classes satisfy the requirements of the axiom of continuity. In like manner if a is any positive number and n is any positive integer, show that there is an x such that $x^n = a$.

21. Theorems on limits and on sets of points. The theorem on limits which is of fundamental algebraic importance is

Theorem 2. If $R(x, y, z, \cdots)$ be any rational function of the variables $x, y, z, \cdots$, and if these variables are approaching limits $X, Y, Z, \cdots$, then the value of R approaches a limit and the limit is $R(X, Y, Z, \cdots)$, provided there is no division by zero.

As any rational expression is made up from its elements by combinations of addition, subtraction, multiplication, and division, it is sufficient to prove the theorem for these four operations. All except the last have been indicated in the above Ex. 3. As multiplication has been cared for, division need be considered only in the simple case of a reciprocal $1/x$. It must be proved that if $\lim x = X$, then $\lim (1/x) = 1/X$. Now

$$\left|\frac{1}{x} - \frac{1}{X}\right| = \frac{|x - X|}{|x|\,|X|}, \qquad \text{by Ex. 3 } (\gamma) \text{ above.}$$

This quantity must be shown to be less than any assigned ϵ. As the quantity is complicated it will be replaced by a simpler one which is greater, owing to an increase in the denominator. Since $x \doteq X$, $x - X$ may be made numerically as small as desired, say less than ϵ', for all x's subsequent to some particular x. Hence if ϵ' be taken at least as small as $\frac{1}{2}|X|$, it appears that $|x|$ must be greater than $\frac{1}{2}|X|$. Then

$$\frac{|x - X|}{|x|\,|X|} < \frac{|x - X|}{\frac{1}{2}|X|^2} < \frac{\epsilon'}{\frac{1}{2}|X|^2}, \qquad \text{by Ex. 3 } (\alpha) \text{ above.}$$

and if ϵ' be restricted to being less than $\frac{1}{2}|X|^2\epsilon$, the difference is less than ϵ and the theorem that $\lim (1/x) = 1/X$ is proved, and also Theorem 2. The necessity for the restriction $X \neq 0$ and the corresponding restriction in the statement of the theorem is obvious.

Theorem 3. If when ϵ is given, no matter how small, it is possible to find a value of n so great that the difference $|x_{n+p} - x_n|$ between x_n and every subsequent term x_{n+p} in the suite $x_1, x_2, \cdots, x_n, \cdots$ is less than ϵ, the suite approaches a limit, and conversely.

The converse part has already been given as Ex. 1 above. The theorem itself is a consequence of the axiom of continuity. First note that as $|x_{n+p} - x_n| < \epsilon$ for all x's subsequent to x_n, the x's cannot become infinite. Suppose 1° that there is some number l such that no matter how remote x_n is in the suite, there are always subsequent values of x which are greater than l and others which are less than l. As all the x's after x_n lie in the interval 2ϵ and as l is less than some x's and greater than others, l must lie in that interval. Hence $|l - x_{n+p}| < 2\epsilon$ for all

x's subsequent to x_n. But now 2ϵ can be made as small as desired because ϵ can be taken as small as desired. Hence the definition of a limit applies and the x's approach l as a limit.

Suppose 2° that there is no such number l. Then every number k is such that either it is possible to go so far in the suite that all subsequent numbers x are as great as k or it is possible to go so far that all subsequent x's are less than k. Hence all numbers k are divided into two classes which satisfy the requirements of the axiom of continuity, and there must be a number N such that the x's ultimately come to lie between $N-\epsilon'$ and $N+\epsilon'$, no matter how small ϵ' is. Hence the x's approach N as a limit. Thus under either supposition the suite approaches a limit and the theorem is proved. It may be noted that under the second supposition the x's ultimately lie entirely upon one side of the point N and that the condition $|x_{n+p}-x_n|<\epsilon$ is not used except to show that the x's remain finite.

22. Consider next a set of points (or their correlative numbers) without any implication that they form a suite, that is, that one may be said to be subsequent to another. If there is only a finite number of points in the set, there is a point farthest to the right and one farthest to the left. If there is an infinity of points in the set, two possibilities arise. Either 1° it is not possible to assign a point K so far to the right that no point of the set is farther to the right — in which case the set is said to be *unlimited above* — or 2° there is a point K such that no point of the set is beyond K — and the set is said to be *limited above.* Similarly, a set may be *limited below* or *unlimited below.* If a set is limited above and below so that it is entirely contained in a finite interval, it is said merely to be *limited.* If there is a point C such that in any interval, no matter how small, surrounding C there are points of the set, then C is called a *point of condensation* of the set (C itself may or may not belong to the set).

Theorem 4. Any infinite set of points which is limited has an upper frontier (maximum ?), a lower frontier (minimum ?), and at least one point of condensation.

Before proving this theorem, consider three infinite sets as illustrations:

(α) 1, 1.9, 1.99, 1.999, $\cdots$, (β) $-2, \cdots, -1.99, -1.9, -1$,

(γ) $-1, -\frac{1}{2}, -\frac{1}{4}, \cdots, \frac{1}{4}, \frac{1}{2}, 1$.

In (α) the element 1 is the minimum and serves also as the lower frontier; it is clearly not a point of condensation, but is isolated. There is no maximum; but 2 is the upper frontier and also a point of condensation. In (β) there is a maximum -1 and a minimum -2 (for -2 has been incorporated with the set). In (γ) there is a maximum and minimum; the point of condensation is 0. If one could be sure that an infinite set had a maximum and minimum, as is the case with finite sets, there would be no need of considering upper and lower frontiers. It is clear that if the upper or lower frontier belongs to the set, there is a maximum or minimum and the frontier is not necessarily a point of condensation; whereas

if the frontier does not belong to the set, it is necessarily a point of condensation and the corresponding extreme point is missing.

To prove that there is an upper frontier, divide the points of the line into two classes, one consisting of points which are to the left of some point of the set, the other of points which are not to the left of any point of the set — then apply the axiom. Similarly for the lower frontier. To show the existence of a point of condensation, note that as there is an infinity of elements in the set, any point p is such that either there is an infinity of points of the set to the right of it or there is not. Hence the two classes into which all points are to be assorted are suggested, and the application of the axiom offers no difficulty.

EXERCISES

1. In a manner analogous to the proof of Theorem 2, show that

$$(\alpha)\ \lim_{x \doteq 0} \frac{x-1}{x-2} = \frac{1}{2}, \qquad (\beta)\ \lim_{x \doteq 2} \frac{3x-1}{x+5} = \frac{5}{7}, \qquad (\gamma)\ \lim_{x \doteq -1} \frac{x^2+1}{x^3-1} = -1.$$

2. Given an infinite series $S = u_1 + u_2 + u_3 + \cdots$. Construct the suite

$$S_1 = u_1,\ S_2 = u_1 + u_2,\ S_3 = u_1 + u_2 + u_3,\ \cdots,\ S_i = u_1 + u_2 + \cdots + u_i,\ \cdots,$$

where S_i is the sum of the first i terms. Show that Theorem 3 gives: The necessary and sufficient condition that the series S converge is that it is possible to find an n so large that $|S_{n+p} - S_n|$ shall be less than an assigned ϵ for all values of p. It is to be understood that a series *converges* when the suite of S's approaches a limit, and conversely.

3. If in a series $u_1 - u_2 + u_3 - u_4 + \cdots$ the terms approach the limit 0, are alternately positive and negative, and each term is less than the preceding, the series converges. Consider the suites $S_1, S_3, S_5, \cdots$ and $S_2, S_4, S_6, \cdots$.

4. Given three infinite suites of numbers

$$x_1, x_2, \cdots, x_n, \cdots; \qquad y_1, y_2, \cdots, y_n, \cdots; \qquad z_1, z_2, \cdots, z_n, \cdots$$

of which the first never decreases, the second never increases, and the terms of the third lie between corresponding terms of the first two, $x_n \leqq z_n \leqq y_n$. Show that the suite of z's has a point of condensation at or between the limits approached by the x's and by the y's; and that if $\lim x = \lim y = l$, then the z's approach l as a limit.

5. Restate the definitions and theorems on sets of points in arithmetic terms.

6. Give the details of the proof of Theorem 4. Show that the proof as outlined gives the least point of condensation. How would the proof be worded so as to give the greatest point of condensation? Show that if a set is limited above, it has an upper frontier but need not have a lower frontier.

7. If a set of points is such that between any two there is a third, the set is said to be *dense*. Show that the rationals form a dense set; also the irrationals. Show that any point of a dense set is a point of condensation for the set.

8. Show that the rationals p/q where $q < K$ do not form a dense set — in fact are a finite set in any limited interval. Hence in regarding any irrational as the limit of a set of rationals it is necessary that the denominators and also the numerators should become infinite.

9. Show that if an infinite set of points lies in a limited region of the plane, say in the rectangle $a \leqq x \leqq b$, $c \leqq y \leqq d$, there must be at least one point of condensation of the set. Give the necessary definitions and apply the axiom of continuity successively to the abscissas and ordinates.

23. Real functions of a real variable. *If x be a variable which takes on a certain set of values of which the totality may be denoted by $[x]$ and if y is a second variable the value of which is uniquely determined for each x of the set $[x]$, then y is said to be a function of x defined over the set $[x]$.* The terms "limited," "unlimited," "limited above," "unlimited below," $\cdots$ are applied to a function if they are applicable to the set $[y]$ of values of the function. Hence Theorem 4 has the corollary:

THEOREM 5. If a function is limited over the set $[x]$, it has an upper frontier M and a lower frontier m for that set.

If the function takes on its upper frontier M, that is, if there is a value x_0 in the set $[x]$ such that $f(x_0) = M$, the function has the absolute *maximum* M at x_0; and similarly with respect to the lower frontier. In any case, the difference $M - m$ between the upper and lower frontiers is called the *oscillation* of the function for the set $[x]$. The set $[x]$ is generally an interval.

Consider some illustrations of functions and sets over which they are defined. The reciprocal $1/x$ is defined for all values of x save 0. In the neighborhood of 0 the function is unlimited above for positive x's and unlimited below for negative x's. It should be noted that the function is not limited in the interval $0 < x \leqq a$ but is limited in the interval $\epsilon \leqq x \leqq a$ where ϵ is any assigned positive number. The function $+\sqrt{x}$ is defined for all positive x's including 0 and is limited below. It is not limited above for the totality of all positive numbers; but if K is assigned, the function is limited in the interval $0 \leqq x \leqq K$. The factorial function $x!$ is defined only for positive integers, is limited below by the value 1, but is not limited above unless the set $[x]$ is limited above. The function $E(x)$ denoting the integer not greater than x or "the integral part of x" is defined for all positive numbers — for instance $E(3) = E(\pi) = 3$. This function is not expressed, like the elementary functions of calculus, as a "formula"; it is defined by a definite law, however, and is just as much of a function as $x^2 + 3x + 2$ or $\frac{1}{2}\sin^2 2x + \log x$. Indeed it should be noted that the elementary functions themselves are in the first instance defined by definite laws and that it is not until after they have been made the subject of considerable study and have been largely developed along analytic lines that they appear as formulas. The ideas of function and formula are essentially distinct and the latter is essentially secondary to the former.

The definition of function as given above excludes the so-called *multiple-valued* functions such as $\sqrt{x}$ and $\sin^{-1} x$ where to a given value of x correspond more than one value of the function. It is usual, however, in treating multiple-valued functions to resolve the functions into different parts or *branches* so that each branch is a single-valued function. Thus $+\sqrt{x}$ is one branch and $-\sqrt{x}$ the other branch

of $\sqrt{x}$; in fact when x is positive the symbol $\sqrt{x}$ is usually restricted to mean merely $+\sqrt{x}$ and thus becomes a single-valued symbol. One branch of $\sin^{-1} x$ consists of the values between $-\frac{1}{2}\pi$ and $+\frac{1}{2}\pi$, other branches give values between $\frac{1}{2}\pi$ and $\frac{3}{2}\pi$ or $-\frac{1}{2}\pi$ and $-\frac{3}{2}\pi$, and so on. Hence the term "function" will be restricted in this chapter to the single-valued functions allowed by the definition.

24. *If $x = a$ is any point of an interval over which $f(x)$ is defined, the function $f(x)$ is said to be continuous at the point $x = a$ if*

$$\lim_{x \doteq a} f(x) = f(a), \qquad \textit{no matter how } x \doteq a.$$

The function is said to be continuous in the interval if it is continuous at every point of the interval. If the function is not continuous at the point a, it is said to be *discontinuous* at a; and if it fails to be continuous at any one point of an interval, it is said to be discontinuous in the interval.

THEOREM 6. If any finite number of functions are continuous (at a point or over an interval), any rational expression formed of those functions is continuous (at the point or over the interval) provided no division by zero is called for.

THEOREM 7. If $y = f(x)$ is continuous at x_0 and takes the value $y_0 = f(x_0)$ and if $z = \phi(y)$ is a continuous function of y at $y = y_0$, then $z = \phi[f(x)]$ will be a continuous function of x at x_0.

In regard to the definition of continuity note that a function cannot be continuous at a point unless it is defined at that point. Thus e^{-1/x^2} is not continuous at $x = 0$ because division by 0 is impossible and the function is undefined. If, however, the function be defined at 0 as $f(0) = 0$, the function becomes continuous at $x = 0$. In like manner the function $1/x$ is not continuous at the origin, and in this case it is impossible to assign to $f(0)$ any value which will render the function continuous; the function becomes infinite at the origin and the very idea of becoming infinite precludes the possibility of approach to a definite limit. Again, the function $E(x)$ is in general continuous, but is discontinuous for integral values of x. When a function is discontinuous at $x = a$, the *amount of the discontinuity* is the limit of the oscillation $M - m$ of the function in the interval $a - \delta < x < a + \delta$ surrounding the point a when δ approaches zero as its limit. The discontinuity of $E(x)$ at each integral value of x is clearly 1; that of $1/x$ at the origin is infinite no matter what value is assigned to $f(0)$.

In case the interval over which $f(x)$ is defined has end points, say $a \leqq x \leqq b$, the question of continuity at $x = a$ must of course be decided by allowing x to approach a from the right-hand side only; and similarly it is a question of left-handed approach to b. In general, if for any reason it is desired to restrict the approach of a variable to its limit to being one-sided, the notations $x \doteq a^+$ and $x \doteq b^-$ respectively are used to denote approach through greater values (right-handed) and through lesser values (left-handed). It is not necessary to make this specification in the case of the ends of an interval; for it is understood that x shall take on only values in the interval in question. It should be noted that

$\lim f(x) = f(x_0)$ when $x \doteq x_0{}^+$ in no wise implies the continuity of $f(x)$ at x_0; a simple example is that of $E(x)$ at the positive integral points.

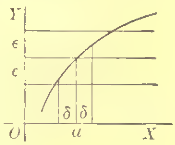
The proof of Theorem 6 is an immediate corollary application of Theorem 2. For

$$\lim R[f(x), \phi(x) \cdots] = R[\lim f(x), \lim \phi(x), \cdots] = R[f(\lim x), \phi(\lim x), \cdots],$$

and the proof of Theorem 7 is equally simple.

Theorem 8. If $f(x)$ is continuous at $x = a$, then for any positive ϵ which has been assigned, no matter how small, there may be found a number δ such that $|f(x) - f(a)| < \epsilon$ in the interval $|x - a| < \delta$, and hence in this interval the oscillation of $f(x)$ is less than 2ϵ. And conversely, if these conditions hold, the function is continuous.

This theorem is in reality nothing but a restatement of the definition of continuity combined with the definition of a limit. For "$\lim f(x) = f(a)$ when $x \doteq a$, no matter how" means that the difference between $f(x)$ and $f(a)$ can be made as small as desired by taking x sufficiently near to a; and conversely. The reason for this restatement is that the present form is more amenable to analytic operations. It also suggests the geometric picture which corresponds to the usual idea of continuity in graphs. For the theorem states that if the two lines $y = f(a) \pm \epsilon$ be drawn, the graph of the function remains between them for at least the short distance δ on each side of $x = a$; and as ϵ may be assigned a value as small as desired, the graph cannot exhibit breaks. On the other hand it should be noted that the actual physical graph is not a curve but a band, a two-dimensional region of greater or less breadth, and that a function could be discontinuous at every point of an interval and yet lie entirely within the limits of any given physical graph.



It is clear that δ, which has to be determined *subsequently* to ϵ, is in general more and more restricted as ϵ is taken smaller and that for different points it is more restricted as the graph rises more rapidly. Thus if $f(x) = 1/x$ and $\epsilon = 1/1000$, δ can be nearly 1/10 if $x_0 = 100$, but must be slightly less than 1/1000 if $x_0 = 1$, and something less than 10^{-6} if x is 10^{-3}. Indeed, if x be allowed to approach zero, the value δ for any assigned ϵ also approaches zero; and although the function $f(x) = 1/x$ is continuous in the interval $0 < x \leqq 1$ and for any given x_0 and ϵ a number δ may be found such that $|f(x) - f(x_0)| < \epsilon$ when $|x - x_0| < \delta$, yet it is not possible to assign a number δ which shall serve *uniformly* for all values of x_0.

25. Theorem 9. If a function $f(x)$ is continuous in an interval $a \leqq x \leqq b$ with end points, it is possible to find a δ such that $|f(x) - f(x_0)| < \epsilon$ when $|x - x_0| < \delta$ for all points x_0; and the function is said to be *uniformly continuous*.

The proof is conducted by the method of reductio ad absurdum. Suppose ϵ is assigned. Consider the suite of values $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots$, or any other suite which approaches zero as a limit. Suppose that no one of these values will serve as a δ for all points of the interval. Then there must be at least one point for which $\frac{1}{2}$ will not serve, at least one for which $\frac{1}{4}$ will not serve, at least one for which $\frac{1}{8}$ will not serve, and so on indefinitely. This infinite set of points must have at least one

point of condensation C such that in any interval surrounding C there are points for which 2^{-k} will not serve as δ, no matter how large k. But now by hypothesis $f(x)$ is continuous at C and hence a number δ can be found such that $|f(x)-f(C)|<\frac{1}{2}\epsilon$ when $|x-x_0|<2\delta$. The oscillation of $f(x)$ in the whole interval 4δ is less than ϵ. Now if x_0 be any point in the middle half of this interval, $|x_0-C|<\delta$; and if x satisfies the relation $|x-x_0|<\delta$, it must still lie in the interval 4δ and the difference $|f(x)-f(x_0)|<\epsilon$, being surely not greater than the oscillation of f in the whole interval. Hence it is possible to surround C with an interval so small that the same δ will serve for any point of the interval. This contradicts the former conclusion, and hence the hypothesis upon which that conclusion was based must have been false and it must have been possible to find a δ which would serve for all points of the interval. The reason why the proof would not apply to a function like $1/x$ defined in the interval $0<x\leqq 1$ lacking an end point is precisely that the point of condensation C would be 0, and at 0 the function is not continuous and $|f(x)-f(C)|<\frac{1}{2}\epsilon$, $|x-C|<2\delta$ could not be satisfied.

Theorem 10. If a function is continuous in a region which includes its end points, the function is limited.

Theorem 11. If a function is continuous in an interval which includes its end points, the function takes on its upper frontier and has a maximum M; similarly it has a minimum m.

These are successive corollaries of Theorem 9. For let ϵ be assigned and let δ be determined so as to serve uniformly for all points of the interval. Divide the interval $b-a$ into n successive intervals of length δ or less. Then in each such interval f cannot increase by more than ϵ nor decrease by more than ϵ. Hence f will be contained between the values $f(a)+n\epsilon$ and $f(a)-n\epsilon$, and is limited. And $f(x)$ has an upper and a lower frontier in the interval. Next consider the rational function $1/(M-f)$ of f. By Theorem 6 this is continuous in the interval unless the denominator vanishes, and if continuous it is limited. This, however, is impossible for the reason that, as M is a frontier of values of f, the difference $M-f$ may be made as small as desired. Hence $1/(M-f)$ is not continuous and there must be some value of x for which $f=M$.

Theorem 12. If $f(x)$ is continuous in the interval $a\leqq x\leqq b$ with end points and if $f(a)$ and $f(b)$ have opposite signs, there is at least one point ξ, $a<\xi<b$, in the interval for which the function vanishes. And whether $f(a)$ and $f(b)$ have opposite signs or not, there is a point ξ, $a<\xi<b$, such that $f(\xi)=\mu$, where μ is any value intermediate between the maximum and minimum of f in the interval.

For convenience suppose that $f(a)<0$. Then in the neighborhood of $x=a$ the function will remain negative on account of its continuity; and in the neighborhood of b it will remain positive. Let ξ be the lower frontier of values of x which make $f(x)$ positive. Suppose that $f(\xi)$ were either positive or negative. Then as f is continuous, an interval could be chosen surrounding ξ and so small that f remained positive or negative in that interval. In neither case could ξ be the lower frontier of positive values. Hence the contradiction, and $f(\xi)$ must be zero. To

prove the second part of the theorem, let c and d be the values of x which make f a minimum and maximum. Then the function $f - \mu$ has opposite signs at c and d, and must vanish at some point of the interval between c and d; and hence a fortiori at some point of the interval from a to b.

EXERCISES

1. Note that x is a continuous function of x, and that consequently it follows from Theorem 6 that any rational fraction $P(x)/Q(x)$, where P and Q are polynomials in x, must be continuous for all x's except roots of $Q(x) = 0$.

2. Graph the function $x - E(x)$ for $x \geqq 0$ and show that it is continuous except for integral values of x. Show that it is limited, has a minimum 0, an upper frontier 1, but no maximum.

3. Suppose that $f(x)$ is defined for an infinite set $[x]$ of which $x = a$ is a point of condensation (not necessarily itself a point of the set). Suppose

$$\lim_{x', x'' \doteq a} [f(x') - f(x'')] = 0 \quad \text{or} \quad |f(x') - f(x'')| < \epsilon, \ |x' - a| < \delta, \ |x'' - a| < \delta,$$

when x' and x'' regarded as *independent* variables approach a as a limit (passing only over values of the set $[x]$, of course). Show that $f(x)$ approaches a limit as $x \doteq a$. By considering the set of values of $f(x)$, the method of Theorem 3 applies almost verbatim. Show that there is no essential change in the proof if it be assumed that x' and x'' become infinite, the set $[x]$ being unlimited instead of having a point of condensation a.

4. From the formula $\sin x < x$ and the formulas for $\sin u - \sin v$ and $\cos u - \cos v$ show that $\Delta \sin x$ and $\Delta \cos x$ are numerically less than $2|\Delta x|$; hence infer that $\sin x$ and $\cos x$ are continuous functions of x for all values of x.

5. What are the intervals of continuity for $\tan x$ and $\csc x$? If $\epsilon = 10^{-4}$, what are approximately the largest available values of δ that will make $|f(x) - f(x_0)| < \epsilon$ when $x_0 = 1°, 30°, 60°, 89°$ for each? Use a four-place table.

6. Let $f(x)$ be defined in the interval from 0 to 1 as equal to 0 when x is irrational and equal to $1/q$ when x is rational and expressed as a fraction p/q in lowest terms. Show that f is continuous for irrational values and discontinuous for rational values. Ex. 8, p. 39, will be of assistance in treating the irrational values.

7. Note that in the definition of continuity a generalization may be introduced by allowing the set $[x]$ over which f is defined to be any set each point of which is a point of condensation of the set, and that hence continuity over a dense set (Ex. 7 above), say the rationals or irrationals, may be defined. This is important because many functions are in the first instance defined only for rationals and are subsequently defined for irrationals by interpolation. Note that if a function is continuous over a dense set (say, the rationals), it does not follow that it is uniformly continuous over the set. For the point of condensation C which was used in the proof of Theorem 9 may not be a point of the set (may be irrational), and the proof would fall through for the same reason that it would in the case of $1/x$ in the interval $0 < x \leqq 1$, namely, because it could not be affirmed that the function was continuous at C. Show that if a function is defined and is uniformly continuous over a dense set, the value $f(x)$ will approach a limit when x approaches any value a (not necessarily of the set, but situated between the upper and lower

frontiers of the set), and that if this limit be defined as the value of $f(a)$, the function will remain continuous. Ex. 3 may be used to advantage.

8. By factoring $(x+\Delta x)^n - x^n$, show for integral values of n that when $0 \leqq x \leqq K$, then $\Delta(x^n) < nK^{n-1}\Delta x$ for small Δx's and consequently x^n is uniformly continuous in the interval $0 \leqq x \leqq K$. If it be assumed that x^n has been defined only for rational x's, it follows from Ex. 7 that the definition may be extended to all x's and that the resulting x^n will be continuous.

9. Suppose (α) that $f(x)+f(y)=f(x+y)$ for any numbers x and y. Show that $f(n)=nf(1)$ and $nf(1/n)=f(1)$, and hence infer that $f(x)=xf(1)=Cx$, where $C=f(1)$, for all rational x's. From Ex. 7 it follows that if $f(x)$ is continuous, $f(x)=Cx$ for all x's. Consider (β) the function $f(x)$ such that $f(x)f(y)=f(x+y)$. Show that it is $Ce^x=a^x$.

10. Show by Theorem 12 that if $y=f(x)$ is a continuous constantly increasing function in the interval $a \leqq x \leqq b$, then to each value of y corresponds a single value of x so that the function $x=f^{-1}(y)$ exists and is single-valued; show also that it is continuous and constantly increasing. State the corresponding theorem if $f(x)$ is constantly decreasing. The function $f^{-1}(y)$ is called the *inverse* function to $f(x)$.

11. Apply Ex. 10 to discuss $y=\sqrt[n]{x}$, where n is integral, x is positive, and only positive roots are taken into consideration.

12. In arithmetic it may readily be shown that the equations

$$a^m a^n = a^{m+n}, \qquad (a^m)^n = a^{mn}, \qquad a^n b^n = (ab)^n,$$

are true when a and b are rational and positive and when m and n are any positive and negative integers or zero. (α) Can it be inferred that they hold when a and b are positive irrationals? (β) How about the extension of the fundamental inequalities

$$x^n > 1, \quad \text{when} \quad x > 1, \qquad x^n < 1, \quad \text{when} \quad 0 \leqq x < 1$$

to all rational values of n and the proof of the inequalities

$$x^m > x^n \quad \text{if} \quad m > n \quad \text{and} \quad x > 1, \qquad x^m < x^n \quad \text{if} \quad m > n \quad \text{and} \quad 0 < x < 1.$$

(γ) Next consider x as held constant and the exponent n as variable. Discuss the exponential function a^x from this relation, and Exs. 10, 11, and other theorems that may seem necessary. Treat the logarithm as the inverse of the exponential.

26. The derivative. *If $x=a$ is a point of an interval over which $f(x)$ is defined and if the quotient*

$$\frac{\Delta f}{\Delta x} = \frac{f(a+h)-f(a)}{h}, \qquad h = \Delta x,$$

approaches a limit when h approaches zero, no matter how, the function $f(x)$ is said to be differentiable at $x=a$ and the value of the limit of the quotient is the derivative $f'(a)$ of f at $x=a$. In the case of differentiability, the definition of a limit gives

$$\frac{f(a+h)-f(a)}{h} = f'(a)+\eta \quad \text{or} \quad f(a+h)-f(a) = hf'(a)+\eta h, \quad (1)$$

where $\lim \eta = 0$ when $\lim h = 0$, no matter how.

In other words if ϵ is given, a δ can be found so that $|\eta| < \epsilon$ when $|h| < \delta$. This shows that a function differentiable at a as in (1) is continuous at a. For

$$|f(a+h) - f(a)| \leqq |f'(a)|\,\delta + \epsilon\delta, \qquad |h| < \delta.$$

If the limit of the quotient exists when $h \doteq 0$ through positive values only, the function has a right-hand derivative which may be denoted by $f'(a^+)$ and similarly for the left-hand derivative $f'(a^-)$. At the end points of an interval the derivative is always considered as one-handed; but for interior points the right-hand and left-hand derivatives must be equal if the function is to have a derivative (unqualified). The function is said to have an *infinite derivative* at a if the quotient becomes infinite as $h \doteq 0$; but if a is an interior point, the quotient must become positively infinite or negatively infinite for all manners of approach and not positively infinite for some and negatively infinite for others. Geometrically this allows a vertical tangent with an inflection point, but not with a cusp as in Fig. 3, p. 8. If infinite derivatives are allowed, the function may have a derivative and yet be discontinuous, as is suggested by any figure where $f(a)$ is any value between $\lim f(x)$ when $x \doteq a^+$ and $\lim f(x)$ when $x \doteq a^-$.

THEOREM 13. If a function takes on its maximum (or minimum) at an interior point of the interval of definition and if it is differentiable at that point, the derivative is zero.

THEOREM 14. *Rolle's Theorem.* If a function $f(x)$ is continuous over an interval $a \leqq x \leqq b$ with end points and vanishes at the ends and has a derivative at each interior point $a < x < b$, there is some point ξ, $a < \xi < b$, such that $f'(\xi) = 0$.

THEOREM 15. *Theorem of the Mean.* If a function is continuous over an interval $a \leqq x \leqq b$ and has a derivative at each interior point, there is some point ξ such that

$$\frac{f(b) - f(a)}{b - a} = f'(\xi) \quad \text{or} \quad \frac{f(a+h) - f(a)}{h} = f'(a + \theta h),$$

where $h \leqq b - a$* and θ is a proper fraction, $0 < \theta < 1$.

To prove the first theorem, note that if $f(a) = M$, the difference $f(a+h) - f(a)$ cannot be positive for any value of h and the quotient $\Delta f/h$ cannot be positive when $h > 0$ and cannot be negative when $h < 0$. Hence the right-hand derivative cannot be positive and the left-hand derivative cannot be negative. As these two must be equal if the function has a derivative, it follows that they must be zero, and the derivative is zero. The second theorem is an immediate corollary. For as the function is continuous it must have a maximum and a minimum (Theorem 11) both of which cannot be zero unless the function is always zero in the interval. Now if the function is identically zero, the derivative is identically zero and the theorem is true; whereas if the function is not identically zero, either the maximum or minimum must be at an interior point, and at that point the derivative will vanish.

* That the theorem is true for any part of the interval from a to b if it is true for the whole interval follows from the fact that the conditions, namely, that f be continuous and that f' exist, hold for any part of the interval if they hold for the whole.

To prove the last theorem construct the auxiliary function

$$\psi(x) = f(x) - f(a) - (x - a)\frac{f(b) - f(a)}{b - a}, \qquad \psi'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

As $\psi(a) = \psi(b) = 0$, Rolle's Theorem shows that there is some point for which $\psi'(\xi) = 0$, and if this value be substituted in the expression for $\psi'(x)$ the solution for $f'(\xi)$ gives the result demanded by the theorem. The proof, however, requires the use of the function $\psi(x)$ and its derivative and is not complete until it is shown that $\psi(x)$ really satisfies the conditions of Rolle's Theorem, namely, is continuous in the interval $a \leqq x \leqq b$ and has a derivative for every point $a < x < b$. The continuity is a consequence of Theorem 6; that the derivative exists follows from the direct application of the definition combined with the assumption that the derivative of f exists.

27. Theorem 16. If a function has a derivative which is identically zero in the interval $a \leqq x \leqq b$, the function is constant; and if two functions have derivatives equal throughout the interval, the functions differ by a constant.

Theorem 17. If $f(x)$ is differentiable and becomes infinite when $x \doteq a$, the derivative cannot remain finite as $x \doteq a$.

Theorem 18. If the derivative $f'(x)$ of a function exists and is a continuous function of x in the interval $a \leqq x \leqq b$, the quotient $\Delta f/h$ converges uniformly toward its limit $f'(x)$.

These theorems are consequences of the Theorem of the Mean. For the first,

$$f(a + h) - f(a) = hf'(a + \theta h) = 0, \quad \text{if} \quad h \leqq b - a, \quad \text{or} \quad f(a + h) = f(a).$$

Hence $f(x)$ is constant. And in case of two functions f and ϕ with equal derivatives, the difference $\psi(x) = f(x) - \phi(x)$ will have a derivative that is zero and the difference will be constant. For the second, let x_0 be a fixed value near a and suppose that in the interval from x_0 to a the derivative remained finite, say less than K. Then

$$|f(x_0 + h) - f(x_0)| = |hf'(x_0 + \theta h)| \leqq h \cdot K.$$

Now let $x_0 + h$ approach a and note that the left-hand term becomes infinite and the supposition that f' remained finite is contradicted. For the third, note that f', being continuous, must be uniformly continuous (Theorem 9), and hence that if ϵ is given, a δ may be found such that

$$\left|\frac{f(x + h) - f(x)}{h} - f'(x)\right| \leqq |f'(x + \theta h) - f'(x)| < \epsilon$$

when $|h| < \delta$ and for all x's in the interval; and the theorem is proved.

Concerning derivatives of higher order no special remarks are necessary. Each is the derivative of a definite function — the previous derivative. If the derivatives of the first n orders exist and are continuous, the derivative of order $n + 1$ may or may not exist. In practical applications, however, the functions are generally indefinitely differentiable except at certain isolated points. The proof of Leibniz's Theorem (§ 8) may be revised so as to depend on elementary processes. Let the formula be assumed for a given value of n. The only terms which can

contribute to the term $D^iuD^{n+1-i}v$ in the formula for the $(n+1)$st derivative of uv are the terms

$$\frac{n(n-1)\cdots(n-i+2)}{1\cdot 2\cdots(i-1)}D^{i-1}u\,D^{n+1-i}v, \qquad \frac{n(n-1)\cdots(n-i+1)}{1\cdot 2\cdots i}D^iu\,D^{n-i}v,$$

in which the first factor is to be differentiated in the first and the second in the second. The sum of the coefficients obtained by differentiating is

$$\frac{n(n-1)\cdots(n-i+2)}{1\cdot 2\cdots(i-1)}+\frac{n(n-1)\cdots(n-i+1)}{1\cdot 2\cdots i}=\frac{(n+1)n\cdots(n-i+2)}{1\cdot 2\cdots i},$$

which is precisely the proper coefficient for the term $D^iuD^{n+1-i}v$ in the expansion of the $(n+1)$st derivative of uv by Leibniz's Theorem.

With regard to this rule and the other elementary rules of operation (4)–(7) of the previous chapter it should be remarked that *a theorem* as well as a rule is involved — thus: If two functions u and v are differentiable at x_0, then the product uv is differentiable at x_0, and the value of the derivative is $u(x_0)v'(x_0)+u'(x_0)v(x_0)$. And similar theorems arise in connection with the other rules. As a matter of fact the ordinary proof needs only to be gone over with care in order to convert it into a rigorous demonstration. But care does need to be exercised both in stating the theorem and in looking to the proof. For instance, the above theorem concerning a product is not true if infinite derivatives are allowed. For let u be -1, 0, or $+1$ according as x is negative, 0, or positive, and let $v=x$. Now v has always a derivative which is 1 and u has always a derivative which is 0, $+\infty$, or 0 according as x is negative, 0, or positive. The product uv is $|x|$, of which the derivative is -1 for negative x's, $+1$ for positive x's, and *nonexistent* for 0. Here the product has no derivative at 0, although each factor has a derivative, and it would be useless to have a formula for attempting to evaluate something that did not exist.

EXERCISES

1. Show that if at a point the derivative of a function exists and is positive, the function must be increasing at that point.

2. Suppose that the derivatives $f'(a)$ and $f'(b)$ exist and are not zero. Show that $f(a)$ and $f(b)$ are relative maxima or minima of f in the interval $a \leqq x \leqq b$, and determine the precise criteria in terms of the signs of the derivatives $f'(a)$ and $f'(b)$.

3. Show that if a continuous function has a positive right-hand derivative at every point of the interval $a \leqq x \leqq b$, then $f(b)$ is the maximum value of f. Similarly, if the right-hand derivative is negative, show that $f(b)$ is the minimum of f.

4. Apply the Theorem of the Mean to show that if $f'(x)$ is continuous at a, then

$$\lim_{x',\,x''\doteq a}\frac{f(x')-f(x'')}{x'-x''}=f'(a),$$

x' and x'' being regarded as independent.

5. Form the increments of a function f for *equicrescent* values of the variable:

$$\Delta_1 f = f(a+h)-f(a), \qquad \Delta_2 f = f(a+2h)-f(a+h),$$
$$\Delta_3 f = f(a+3h)-f(a+2h), \cdots.$$

These are called first differences ; the differences of these differences are

$$\Delta_1^2 f = f(a + 2h) - 2f(a + h) + f(a),$$
$$\Delta_2^2 f = f(a + 3h) - 2f(a + 2h) + f(a + h), \cdots$$

which are called the second differences ; in like manner there are third differences

$$\Delta_1^3 f = f(a + 3h) - 3f(a + 2h) + 3f(a + h) - f(a), \cdots$$

and so on. Apply the Law of the Mean to all the differences and show that

$$\Delta_1^2 f = h^2 f''(a + \theta_1 h + \theta_2 h), \qquad \Delta_1^3 f = h^3 f'''(a + \theta_1 h + \theta_2 h + \theta_3 h), \cdots.$$

Hence show that if the first n derivatives of f are continuous at a, then

$$f''(a) = \lim_{h \doteq 0} \frac{\Delta^2 f}{h^2}, \qquad f'''(a) = \lim_{h \doteq 0} \frac{\Delta^3 f}{h^3}, \qquad \cdots, \qquad f^{(n)}(a) = \lim_{h \doteq 0} \frac{\Delta^n f}{h^n}.$$

6. *Cauchy's Theorem.* If $f(x)$ and $\phi(x)$ are continuous over $a \leqq x \leqq b$, have derivatives at each interior point, and if $\phi'(x)$ does not vanish in the interval,

$$\frac{f(b) - f(a)}{\phi(b) - \phi(a)} = \frac{f'(\xi)}{\phi'(\xi)} \quad \text{or} \quad \frac{f(a + h) - f(a)}{\phi(a + h) - \phi(a)} = \frac{f'(a + \theta h)}{\phi'(a + \theta h)}.$$

Prove that this follows from the application of Rolle's Theorem to the function

$$\psi(x) = f(x) - f(a) - [\phi(x) - \phi(a)]\frac{f(b) - f(a)}{\phi(b) - \phi(a)}.$$

7. One application of Ex. 6 is to the theory of indeterminate forms. Show that if $f(a) = \phi(a) = 0$ and if $f'(x)/\phi'(x)$ approaches a limit when $x \doteq a$, then $f(x)/\phi(x)$ will approach the same limit.

8. *Taylor's Theorem.* Note that the form $f(b) = f(a) + (b - a)f'(\xi)$ is one way of writing the Theorem of the Mean. By the application of Rolle's Theorem to

$$\psi(x) = f(b) - f(x) - (b - x)f'(x) - (b - x)^2\frac{f(b) - f(a) - (b - a)f'(a)}{(b - a)^2},$$

show $$f(b) = f(a) + (b - a)f'(a) + \frac{(b - a)^2}{2}f''(\xi),$$

and to $$\psi(x) = f(b) - f(x) - (b - x)f'(x) - \frac{(b - x)^2}{2}f''(x) - \cdots - \frac{(b - x)^{n-1}}{(n - 1)!}f^{(n-1)}(x)$$
$$- \frac{(b - x)^n}{(b - a)^n}\Big[f(b) - f(a) - (b - a)f'(a)$$
$$- \frac{(b - a)^2}{2}f''(a) - \cdots - \frac{(b - a)^{n-1}}{(n - 1)!}f^{(n-1)}(a)\Big],$$

show $$f(b) = f(a) + (b - a)f'(a) + \frac{(b - a)^2}{2}f''(a) + \cdots$$
$$+ \frac{(b - a)^{n-1}}{(n - 1)!}f^{(n-1)}(a) + \frac{(b - a)^n}{n!}f^{(n)}(\xi).$$

What are the restrictions that must be imposed on the function and its derivatives?

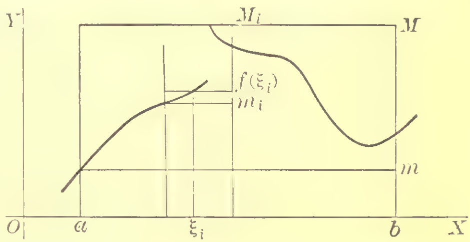
9. If a continuous function over $a \leqq x \leqq b$ has a right-hand derivative at each point of the interval which is zero, show that the function is constant. Apply Ex. 2 to the functions $f(x) + \epsilon(x - a)$ and $f(x) - \epsilon(x - a)$ to show that the maximum difference between the functions is $2\epsilon(b - a)$ and that f must therefore be constant.

10. State and prove the theorems implied in the formulas (4)–(6), p. 2.

11. Consider the extension of Ex. 7, p. 44, to derivatives of functions defined over a dense set. If the derivative exists and is uniformly continuous over the dense set, what of the existence and continuity of the derivative of the function when its definition is extended as there indicated ?

12. If $f(x)$ has a finite derivative at each point of the interval $a \leqq x \leqq b$, the derivative $f'(x)$ must take on every value intermediate between any two of its values. To show this, take first the case where $f'(a)$ and $f'(b)$ have opposite signs and show, by the continuity of f and by Theorem 13 and Ex. 2, that $f'(\xi) = 0$. Next if $f'(a) < \mu < f'(b)$ without any restrictions on $f'(a)$ and $f'(b)$, consider the function $f(x) - \mu x$ and its derivative $f'(x) - \mu$. Finally, prove the complete theorem. It should be noted that the continuity of $f'(x)$ is not assumed, nor is it proved; for there are functions which take every value intermediate between two given values and yet are not continuous.

28. Summation and integration. Let $f(x)$ be defined and limited over the interval $a \leqq x \leqq b$ and let M, m, and $O = M - m$ be the upper frontier, lower frontier, and oscillation of $f(x)$ in the interval. Let $n - 1$ points of division be introduced in the interval dividing it into n consecutive intervals $\delta_1, \delta_2, \cdots, \delta_n$ of which the largest has the length Δ and let M_i, m_i, O_i, and $f(\xi_i)$ be the upper and lower frontiers, the oscillation, and any value of the function in the interval δ_i. Then the inequalities



$$m\delta_i \leqq m_i\delta_i \leqq f(\xi_i)\,\delta_i \leqq M_i\delta_i \leqq M\,\delta_i$$

will hold, and if these terms be summed up for all n intervals,

$$m\,(b - a) \leqq \sum m_i\delta_i \leqq \sum f(\xi_i)\,\delta_i \leqq \sum M_i\delta_i \leqq M\,(b - a) \qquad (A)$$

will also hold. Let $s = \Sigma m_i\delta_i$, $\sigma = \Sigma f(\xi_i)\,\delta_i$, and $S = \Sigma M_i\delta_i$. From (A) it is clear that the difference $S - s$ does not exceed

$$(M - m)\,(b - a) = O\,(b - a),$$

the product of the length of the interval by the oscillation in it. The values of the sums S, s, σ will evidently depend on the number of parts into which the interval is divided and on the way in which it is divided into that number of parts.

Theorem 19. If n' additional points of division be introduced into the interval, the sum S' constructed for the $n + n' - 1$ points of division

cannot be greater than S and cannot be less than S by more than $n'O\Delta$. Similarly, s' cannot be less than s and cannot exceed s by more than $n'O\Delta$.

THEOREM 20. There exists a lower frontier L for all possible methods of constructing the sum S and an upper frontier l for s.

THEOREM 21. *Darboux's Theorem.* When ϵ is assigned it is possible to find a Δ so small that for all methods of division for which $\delta_i \leqq \Delta$, the sums S and s shall differ from their frontier values L and l by less than any preassigned ϵ.

To prove the first theorem note that although (A) is written for the whole interval from a to b and for the sums constructed on it, yet it applies equally to any part of the interval and to the sums constructed on that part. Hence if $S_i = M_i\delta_i$ be the part of S due to the interval δ_i and if S_i' be the part of S' due to this interval after the introduction of some of the additional points into it, $m_i\delta_i \leqq S_i' \leqq S_i = M_i\delta_i$. Hence S_i' is not greater than S_i (and as this is true for each interval δ_i, S' is not greater than S) and, moreover, $S_i - S_i'$ is not greater than $O_i\delta_i$ and a fortiori not greater than $O\Delta$. As there are only n' new points, not more than n' of the intervals δ_i can be affected, and hence the total decrease $S - S'$ in S cannot be more than $n'O\Delta$. The treatment of s is analogous.

Inasmuch as (A) shows that the sums S and s are limited, it follows from Theorem 4 that they possess the frontiers required in Theorem 20. To prove Theorem 21 note first that as L is a frontier for all the sums S, there is some particular sum S which differs from L by as little as desired, say $\frac{1}{2}\epsilon$. For this S let n be the number of divisions. Now consider S' as any sum for which each δ_i is less than $\Delta = \frac{1}{2}\epsilon/nO$. If the sum S'' be constructed by adding the n points of division for S to the points of division for S', S'' cannot be greater than S and hence cannot differ from L by so much as $\frac{1}{2}\epsilon$. Also S'' cannot be greater than S' and cannot be less than S' by more than $nO\Delta$, which is $\frac{1}{2}\epsilon$. As S'' differs from L by less than $\frac{1}{2}\epsilon$ and S' differs from S'' by less than $\frac{1}{2}\epsilon$, S' cannot differ from L by more than ϵ, which was to be proved. The treatment of s and l is analogous.

29. If indices are introduced to indicate the interval for which the frontiers L and l are calculated and if β lies in the interval from a to b, then L_a^β and l_a^β will be functions of β.

THEOREM 22. The equations $L_a^b = L_a^c + L_c^b$, $a < c < b$; $L_a^b = -L_b^a$; $L_a^b = \mu(b-a)$, $m \leqq \mu \leqq M$, hold for L, and similar equations for l. As functions of β, L_a^β and l_a^β are continuous, and if $f(x)$ is continuous, they are differentiable and have the common derivative $f(\beta)$.

To prove that $L_a^b = L_a^c + L_c^b$, consider c as one of the points of division of the interval from a to b. Then the sums S will satisfy $S_a^b = S_a^c + S_c^b$, and as the limit of a sum is the sum of the limits, the corresponding relation must hold for the frontier L. To show that $L_a^b = -L_b^a$ it is merely necessary to note that $S_a^b = -S_b^a$ because in passing from b to a the intervals δ_i must be taken with the sign opposite to that which they have when the direction is from a to b. From (A) it appears that $m(b-a) \leqq S_a^b \leqq M(b-a)$ and hence in the limit $m(b-a) \leqq L_a^b \leqq M(b-a)$.

Hence there is a value μ, $m \leqq \mu \leqq M$, such that $L_a^b = \mu(b-a)$. To show that L_a^β is a continuous function of β, take $K > |M|$ and $|m|$, and consider the relations

$$L_a^{\beta+h} - L_a^\beta = L_a^\beta + L_\beta^{\beta+h} - L_a^\beta = L_\beta^{\beta+h} = \mu h, \qquad |\mu| < K,$$

$$L_a^{\beta-h} - L_a^\beta = L_a^{\beta-h} - L_a^{\beta-h} - L_{\beta-h}^\beta = - L_{\beta-h}^\beta = - \mu' h, \qquad |\mu'| < K.$$

Hence if ϵ is assigned, a δ may be found, namely $\delta < \epsilon/K$, so that $|L_a^{\beta\pm h} - L_a^\beta| < \epsilon$ when $h < \delta$ and L_a^β is therefore continuous. Finally consider the quotients

$$\frac{L_a^{\beta+h} - L_a^\beta}{h} = \mu \quad \text{and} \quad \frac{L_a^{\beta-h} - L_a^\beta}{-h} = \mu',$$

where μ is some number between the maximum and minimum of $f(x)$ in the interval $\beta \leqq x \leqq \beta + h$ and, if f is continuous, is some value $f(\xi)$ of f in that interval and where $\mu' = f(\xi')$ is some value of f in the interval $\beta - h \leqq x \leqq \beta$. Now let $h \doteq 0$. As the function f is continuous, $\lim f(\xi) = f(\beta)$ and $\lim f(\xi') = f(\beta)$. Hence the right-hand and left-hand derivatives exist and are equal and the function L_a^β has the derivative $f(\beta)$. The treatment of l is analogous.

Theorem 23. For a given interval and function f, the quantities l and L satisfy the relation $l \leqq L$; and the necessary and sufficient condition that $L = l$ is that there shall be some division of the interval which shall make $\Sigma(M_i - m_i)\delta_i = \Sigma O_i\delta_i < \epsilon$.

If $L_a^b = l_a^b$, the function f is said to be integrable over the interval from a to b and the integral $\int_a^b f(x)\,dx$ is defined as the common value $L_a^b = l_a^b$. Thus the definite integral is defined.

Theorem 24. If a function is integrable over an interval, it is integrable over any part of the interval and the equations

$$\int_a^c f(x)\,dx + \int_c^b f(x)\,dx = \int_a^b f(x)\,dx,$$

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx, \qquad \int_a^b f(x)\,dx = \mu(b-a)$$

hold; moreover, $\int_a^\beta f(x)\,dx = F(\beta)$ is a continuous function of β; and if $f(x)$ is continuous, the derivative $F'(\beta)$ will exist and be $f(\beta)$.

By (A) the sums S and s constructed for the same division of the interval satisfy the relation $S - s \geqq 0$. By Darboux's Theorem the sums S and s will approach the values L and l when the divisions are indefinitely decreased. Hence $L - l \geqq 0$. Now if $L = l$ and a Δ be found so that when $\delta_i < \Delta$ the inequalities $S - L < \frac{1}{2}\epsilon$ and $l - s < \frac{1}{2}\epsilon$ hold, then $S - s = \Sigma(M_i - m_i)\delta_i = \Sigma O_i\delta_i < \epsilon$; and hence the condition $\Sigma O_i\delta_i < \epsilon$ is seen to be necessary. Conversely if there is any method of division such that $\Sigma O_i\delta_i < \epsilon$, then $S - s < \epsilon$ and the lesser quantity $L - l$ must also be less than ϵ. But if the difference between two constant quantities can be made less than ϵ, where ϵ is arbitrarily assigned, the constant quantities are equal; and hence the

condition is seen to be also sufficient. To show that if a function is integrable over an interval, it is integrable over any part of the interval, it is merely necessary to show that if $L_a^b = l_a^b$, then $L_\alpha^\beta = l_\alpha^\beta$ where α and β are two points of the interval. Here the condition $\Sigma O_i\delta_i < \epsilon$ applies; for if $\Sigma O_i\delta_i$ can be made less than ϵ for the whole interval, its value for any part of the interval, being less than for the whole, must be less than ϵ. The rest of Theorem 24 is a corollary of Theorem 22.

30. Theorem 25. A function is integrable over the interval $a \leqq x \leqq b$ if it is continuous in that interval.

Theorem 26. If the interval $a \leqq x \leqq b$ over which $f(x)$ is defined and limited contains only a finite number of points at which f is discontinuous or if it contains an infinite number of points at which f is discontinuous but these points have only a finite number of points of condensation, the function is integrable.

Theorem 27. If $f(x)$ is integrable over the interval $a \leqq x \leqq b$, the sum $\sigma = \Sigma f(\xi_i)\delta_i$ will approach the limit $\int_a^b f(x)dx$ when the individual intervals δ_i approach the limit zero, it being immaterial how they approach that limit or how the points ξ_i are selected in their respective intervals δ_i.

Theorem 28. If $f(x)$ is continuous in an interval $a \leqq x \leqq b$, then $f(x)$ has an indefinite integral, namely $\int_a^x f(x)dx$, in the interval.

Theorem 25 may be reduced to Theorem 23. For as the function is continuous, it is possible to find a Δ so small that the oscillation of the function in any interval of length Δ shall be as small as desired (Theorem 9). Suppose Δ be chosen so that the oscillation is less than $\epsilon/(b-a)$. Then $\Sigma O_i\delta_i < \epsilon$ when $\delta_i < \Delta$; and the function is integrable. To prove Theorem 26, take first the case of a finite number of discontinuities. Cut out the discontinuities surrounding each value of x at which f is discontinuous by an interval of length δ. As the oscillation in each of these intervals is not greater than O, the contribution of these intervals to the sum $\Sigma O_i\delta_i$ is not greater than $On\delta$, where n is the number of the discontinuities. By taking δ small enough this may be made as small as desired, say less than $\frac{1}{2}\epsilon$. Now in each of the remaining parts of the interval $a \leqq x \leqq b$, the function f is continuous and hence integrable, and consequently the value of $\Sigma O_i\delta_i$ for these portions may be made as small as desired, say $\frac{1}{2}\epsilon$. Thus the sum $\Sigma O_i\delta_i$ for the whole interval can be made as small as desired and $f(x)$ is integrable. When there are points of condensation they may be treated just as the isolated points of discontinuity were treated. After they have been surrounded by intervals, there will remain over only a finite number of discontinuities. Further details will be left to the reader.

For the proof of Theorem 27, appeal may be taken to the fundamental relation (A) which shows that $s \leqq \sigma \leqq S$. Now let the number of divisions increase indefinitely and each division become indefinitely small. As the function is integrable, S and s approach the same limit $\int_a^b f(x)dx$, and consequently σ which is included between them must approach that limit. Theorem 28 is a corollary of Theorem 24

which states that as $f(x)$ is continuous, the derivative of $\int_a^x f(x)\,dx$ is $f(x)$. By definition, the indefinite integral is any function whose derivative is the integrand. Hence $\int_a^x f(x)\,dx$ is an indefinite integral of $f(x)$, and any other may be obtained by adding to this an arbitrary constant (Theorem 16). Thus it is seen that the proof of the existence of the indefinite integral for any given continuous function is made to depend on the theory of definite integrals.

EXERCISES

1. Rework some of the proofs in the text with l replacing L.

2. Show that the L obtained from $Cf(x)$, where C is a constant, is C times the L obtained from f. Also if u, v, w are all limited in the interval $a \leqq x \leqq b$, the L for the combination $u + v - w$ will be $L(u) + L(v) - L(w)$, where $L(u)$ denotes the L for u, etc. State and prove the corresponding theorems for definite integrals and hence the corresponding theorems for indefinite integrals.

3. Show that $\Sigma O_i \delta_i$ can be made less than an assigned ϵ in the case of the function of Ex. 6, p. 44. Note that $l = 0$, and hence infer that the function is integrable and the integral is zero. The proof may be made to depend on the fact that there are only a finite number of values of the function greater than any assigned value.

4. State with care and prove the results of Exs. 3 and 5, p. 29. What restriction is to be placed on $f(x)$ if $f(\xi)$ may replace μ?

5. State with care and prove the results of Ex. 4, p. 29, and Ex. 13, p. 30.

6. If a function is limited in the interval $a \leqq x \leqq b$ and never decreases, show that the function is integrable. This follows from the fact that $\Sigma O_i \leqq O$ is finite.

7. More generally, let $f(x)$ be such a function that ΣO_i remains less than some number K, no matter how the interval be divided. Show that f is integrable. Such a function is called a *function of limited variation* (§ 127).

8. *Change of variable.* Let $f(x)$ be continuous over $a \leqq x \leqq b$. Change the variable to $x = \phi(t)$, where it is supposed that $a = \phi(t_1)$ and $b = \phi(t_2)$, and that $\phi(t)$, $\phi'(t)$, and $f[\phi(t)]$ are continuous in t over $t_1 \leqq t \leqq t_2$. Show that

$$\int_a^b f(x)\,dx = \int_{t_1}^{t_2} f[\phi(t)]\,\phi'(t)\,dt \quad \text{or} \quad \int_{\phi(t_1)}^{\phi(t)} f(x)\,dx = \int_{t_1}^{t} f[\phi(t)]\,\phi'(t)\,dt.$$

Do this by showing that the derivatives of the two sides of the last equation with respect to t exist and are equal over $t_1 \leqq t \leqq t_2$, that the two sides vanish when $t = t_1$ and are equal, and hence that they must be equal throughout the interval.

9. *Osgood's Theorem.* Let α_i be a set of quantities which differ uniformly from $f(\xi_i)\,\delta_i$ by an amount $\zeta_i \delta_i$, that is, suppose

$$\alpha_i = f(\xi_i)\,\delta_i + \zeta_i \delta_i, \quad \text{where} \quad |\zeta_i| < \epsilon \quad \text{and} \quad a \leqq \xi \leqq b.$$

Prove that if f is integrable, the sum $\Sigma \alpha_i$ approaches a limit when $\delta_i \doteq 0$ and that the limit of the sum is $\int_a^b f(x)\,dx$.

10. Apply Ex. 9 to the case $\Delta f = f'\Delta x + \zeta \Delta x$ where f' is continuous to show directly that $f(b) - f(a) = \int_a^b f'(x)\,dx$. Also by regarding $\Delta x = \phi'(t)\,\Delta t + \zeta \Delta t$, apply to Ex. 8 to prove the rule for change of variable.

PART I. DIFFERENTIAL CALCULUS

CHAPTER III

TAYLOR'S FORMULA AND ALLIED TOPICS

31. Taylor's Formula. The object of Taylor's Formula is to express the value of a function $f(x)$ in terms of the values of the function and its derivatives at some one point $x = a$. Thus

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + R. \qquad (1)$$

Such an expansion is necessarily true because the remainder R may be considered as defined by the equation; the real significance of the formula must therefore lie in the possibility of finding a simple expression for R, and there are several.

THEOREM. On the hypothesis that $f(x)$ and its first n derivatives exist and are continuous over the interval $a \leqq x \leqq b$, the function may be expanded in that interval into a polynomial in $x - a$,

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + R, \qquad (1)$$

with the remainder R expressible in any one of the forms

$$R = \frac{(x-a)^n}{n!}f^{(n)}(\xi) = \frac{h^n(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(\xi) = \frac{1}{(n-1)!}\int_0^h t^{n-1}f^{(n)}(a+h-t)\,dt, \qquad (2)$$

where $h = x - a$ and $a < \xi < x$ or $\xi = a + \theta h$ where $0 < \theta < 1$.

A first proof may be made to depend on Rolle's Theorem as indicated in Ex. 8, p. 49. Let x be regarded for the moment as constant, say equal to b. Construct

the function $\psi(x)$ there indicated. Note that $\psi(a) = \psi(b) = 0$ and that the derivative $\psi'(x)$ is merely

$$\psi'(x) = -\frac{(b-x)^{n-1}}{(n-1)!}f^{(n)}(x) + n\frac{(b-x)^{n-1}}{(b-a)^n}\Big[f(b) - f(a) - (b-a)f'(a) - \cdots - \frac{(b-a)^{n-1}}{(n-1)!}f^{(n-1)}(a)\Big].$$

By Rolle's Theorem $\psi'(\xi) = 0$. Hence if ξ be substituted above, the result is

$$f(b) = f(a) + (b-a)f'(a) + \cdots + \frac{(b-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{(b-a)^n}{n!}f^{(n)}(\xi),$$

after striking out the factor $-(b-\xi)^{n-1}$, multiplying by $(b-a)^n/n$, and transposing $f(b)$. The theorem is therefore proved with the first form of the remainder. *This proof does not require the continuity of the nth derivative nor its existence at a and at b.*

The second form of the remainder may be found by applying Rolle's Theorem to

$$\psi(x) = f(b) - f(x) - (b-x)f'(x) - \cdots - \frac{(b-x)^{n-1}}{(n-1)!}f^{(n-1)}(x) - (b-x)P,$$

where P is determined so that $R = (b-a)P$. Note that $\psi(b) = 0$ and that by Taylor's Formula $\psi(a) = 0$. Now

$$\psi'(x) = -\frac{(b-x)^{n-1}}{(n-1)!}f^{(n)}(x) + P \quad \text{or} \quad P = f^{(n)}(\xi)\frac{(b-\xi)^{n-1}}{(n-1)!} \quad \text{since} \quad \psi'(\xi) = 0.$$

Hence if ξ be written $\xi = a + \theta h$ where $h = b - a$, then $b - \xi = b - a - \theta h = (b-a)(1-\theta)$.

And $$R = (b-a)P = (b-a)\frac{(b-a)^{n-1}(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(\xi) = \frac{(b-a)^n(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(\xi).$$

The second form of R is thus found. In this work as before, the result is proved for $x = b$, the end point of the interval $a \leqq x \leqq b$. But as the interval could be considered as terminating at any of its points, the proof clearly applies to any x in the interval.

A second proof of Taylor's Formula, and the easiest to remember, consists in integrating the nth derivative n times from a to x. The successive results are

$$\int_a^x f^{(n)}(x)\,dx = f^{n-1}(x)\Big]_a^x = f^{(n-1)}(x) - f^{(n-1)}(a).$$

$$\int_a^x\int_a^x f^{(n)}(x)\,dx^2 = \int_a^x f^{(n-1)}(x)\,dx - \int_a^x f^{(n-1)}(a)\,dx$$
$$= f^{(n-2)}(x) - f^{(n-2)}(a) - (x-a)f^{(n-1)}(a).$$

$$\int_a^x\int_a^x\int_a^x f^{(n)}(x)\,dx^3 = f^{(n-3)}(x) - f^{(n-3)}(a) - (x-a)f^{(n-2)}(a) - \frac{(x-a)^2}{2!}f^{(n-1)}(a).$$

$$\int_a^x\cdots\int_a^x f^{(n)}(x)\,dx^n = f(x) - f(a) - (x-a)f'(a) - \frac{(x-a)^2}{2!}f''(a) - \cdots - \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a).$$

The formula is therefore proved with R in the form $\int_a^x\cdots\int_a^x f^{(n)}(x)\,dx^n$. To transform this to the ordinary form, the Law of the Mean may be applied ((65), § 16). For

$$m(x-a) < \int_a^x f^{(n)}(x)\,dx < M(x-a), \qquad m\frac{(x-a)^n}{n!} < \int_a^x\cdots\int_a^x f^{(n)}(x)\,dx^n < M\frac{(x-a)^n}{n!},$$

where m is the least and M the greatest value of $f^{(n)}(x)$ from a to x. There is then some intermediate value $f^{(n)}(\xi)=\mu$ such that

$$\int_a^x\cdots\int_a^x f^{(n)}(x)\,dx^n=\frac{(x-a)^n}{n!}f^{(n)}(\xi).$$

This proof requires that the nth derivative be continuous and is less general.

The third proof is obtained by applying successive integrations by parts to the obvious identity $f(a+h)-f(a)=\int_0^h f'(a+h-t)\,dt$ to make the integrand contain higher derivatives.

$$f(a+h)-f(a)=\int_0^h f'(a+h-t)\,dt=tf'(a+h-t)\Big]_0^h+\int_0^h tf''(a+h-t)\,dt$$

$$=hf'(a)+\tfrac{1}{2}t^2f''(a+h-t)\Big]_0^h+\int_0^h \tfrac{1}{2}t^2f'''(a+h-t)\,dt$$

$$=hf'(a)+\frac{h^2}{2!}f''(a)+\cdots+\frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a)+\int_0^h\frac{t^{n-1}}{(n-1)!}f^{(n)}(a+h-t)\,dt.$$

This, however, is precisely Taylor's Formula with the third form of remainder.

If the point a about which the function is expanded is $x=0$, the expansion will take the form known as Maclaurin's Formula:

$$f(x)=f(0)+xf'(0)+\frac{x^2}{2!}f''(0)+\cdots+\frac{x^{n-1}}{(n-1)!}f^{(n-1)}(0)+R, \quad (3)$$

$$R=\frac{x^n}{n!}f^{(n)}(\theta x)=\frac{x^n}{(n-1)!}(1-\theta)^{n-1}f^{(n)}(\theta x)=\frac{1}{(n-1)!}\int_0^x t^{n-1}f^{(n)}(x-t)\,dt.$$

32. Both Taylor's Formula and its special case, Maclaurin's, express a function as a polynomial in $h=x-a$, of which all the coefficients except the last are constants while the last is not constant but depends on h both explicitly and through the unknown fraction θ which itself is a function of h. If, however, the nth derivative is continuous, the coefficient $f^{(n)}(a+\theta h)/n!$ must remain finite, and if the form of the derivative is known, it may be possible actually to assign limits between which $f^{(n)}(a+\theta h)/n!$ lies. This is of great importance in making approximate calculations as in Exs. 8 ff. below; for it sets a limit to the value of R for any value of n.

Theorem. There is only one possible expansion of a function into a polynomial in $h=x-a$ of which all the coefficients except the last are constant and the last finite; and hence if such an expansion is found in any manner, it must be Taylor's (or Maclaurin's).

To prove this theorem consider two polynomials of the nth order

$$c_0+c_1h+c_2h^2+\cdots+c_{n-1}h^{n-1}+c_nh^n=C_0+C_1h+C_2h^2+\cdots+C_{n-1}h^{n-1}+C_nh^n,$$

which represent the same function and hence are equal for all values of h from 0 to $b-a$. It follows that the coefficients must be equal. For let h approach 0.

The terms containing h will approach 0 and hence c_0 and C_0 may be made as nearly equal as desired; and as they are constants, they must be equal. Strike them out from the equation and divide by h. The new equation must hold for all values of h from 0 to $b - a$ with the possible exception of 0. Again let $h \doteq 0$ and now it follows that $c_1 = C_1$. And so on, with all the coefficients. The two developments are seen to be identical, and hence identical with Taylor's.

To illustrate the application of the theorem, let it be required to find the expansion of $\tan x$ about 0 when the expansions of $\sin x$ and $\cos x$ about 0 are given.

$$\sin x = x - \tfrac{1}{6}x^3 + \tfrac{1}{120}x^5 + Px^7, \qquad \cos x = 1 - \tfrac{1}{2}x^2 + \tfrac{1}{24}x^4 + Qx^6,$$

where P and Q remain finite in the neighborhood of $x = 0$. In the first place note that $\tan x$ clearly has an expansion; for the function and its derivatives (which are combinations of $\tan x$ and $\sec x$) are finite and continuous until x approaches $\frac{1}{2}\pi$. By division,

$$\begin{array}{r|llll}
 & x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 & & & \\
\hline
1 - \tfrac{1}{2}x^2 + \tfrac{1}{24}x^4 + Qx^6 & x - \tfrac{1}{6}x^3 + \tfrac{1}{120}x^5 & \vdots & + Px^7 & \\
 & x - \tfrac{1}{2}x^3 + \tfrac{1}{24}x^5 & \vdots & + Qx^7 & \\
\cline{2-4}
 & \qquad \tfrac{1}{3}x^3 - \tfrac{1}{30}x^5 & \vdots & + (P - Q)x^7 & \\
 & \qquad \tfrac{1}{3}x^3 - \tfrac{1}{6}x^5 & \vdots & + \tfrac{1}{72}x^7 + \tfrac{1}{3}Qx^9 & \\
\cline{2-5}
 & \qquad\qquad \tfrac{2}{15}x^5 & \vdots & \ldots\ldots\ldots\ldots &
\end{array}$$

Hence $\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \dfrac{S'}{\cos x}x^7$, where S is the remainder in the division and is an expression containing P, Q, and powers of x; it must remain finite if P and Q remain finite. The quotient $S/\cos x$ which is the coefficient of x^7 therefore remains finite near $x = 0$, and the expression for $\tan x$ is the Maclaurin expansion up to terms of the sixth order, plus a remainder.

In the case of functions compounded from simple functions of which the expansion is known, this method of obtaining the expansion by algebraic processes upon the known expansions treated as polynomials is generally shorter than to obtain the result by differentiation. The computation may be abridged by omitting the last terms and work such as follows the dotted line in the example above; but if this is done, care must be exercised against carrying the algebraic operations too far or not far enough. In Ex. 5 below, the last terms should be put in and carried far enough to insure that the desired expansion has neither more nor fewer terms than the circumstances warrant.

EXERCISES

1. Assume $R = (b - a)^k P$; show $R = \dfrac{h^n(1 - \theta)^{n-k}}{(n-1)!\,k} f^{(n)}(\xi)$.

2. Apply Ex. 5, p. 29, to compare the third form of remainder with the first.

3. Obtain, by differentiation and substitution in (1), three nonvanishing terms:

(α) $\sin^{-1} x,\ a = 0$. (β) $\tanh x,\ a = 0$. (γ) $\tan x,\ a = \frac{1}{4}\pi$.

(δ) $\csc x,\ a = \frac{1}{4}\pi$. (ε) $e^{\sin x},\ a = 0$. (ζ) $\log \sin x,\ a = \frac{1}{2}\pi$.

4. Find the nth derivatives in the following cases and write the expansion:

(α) $\sin x,\ a = 0$. (β) $\sin x,\ a = \frac{1}{2}\pi$. (γ) $e^x,\ a = 0$.

(δ) $e^x,\ a = 1$. (ε) $\log x,\ a = 1$, (ζ) $(1 + x)^k,\ a = 0$.

5. By algebraic processes find the Maclaurin expansion to the term in x^5:

(α) $\sec x$, $\quad(\beta)$ $\tanh x$, $\quad(\gamma)$ $-\sqrt{1-x^2}$,

(δ) $e^x \sin x$, $\quad(\epsilon)$ $[\log(1-x)]^2$, $\quad(\zeta)$ $+\sqrt{\cosh x}$,

(η) $e^{\sin x}$, $\quad(\theta)$ $\log\cos x$, $\quad(\iota)$ $\log\sqrt{1+x^2}$.

The expansions needed in this work may be found by differentiation or taken from B. O. Peirce's "Tables." In (γ) and (ζ) apply the binomial theorem of Ex. 4 (ζ). In (η) let $y = \sin x$, expand e^y, and substitute for y the expansion of $\sin x$. In (θ) let $\cos x = 1 - y$. In all cases show that the coefficient of the term in x^6 really remains finite when $x \doteq 0$.

6. If $f(a+h) = c_0 + c_1 h + c_2 h^2 + \cdots + c_{n-1}h^{n-1} + c_n h^n$, show that in

$$\int_0^h f(a+h)\,dh = c_0 h + \frac{c_1}{2}h^2 + \frac{c_2}{3}h^3 + \cdots + \frac{c_{n-1}}{n}h^n + \int_0^h c_n h^n dh$$

the last term may really be put in the form Ph^{n+1} with P finite. Apply Ex. 5, p. 29.

7. Apply Ex. 6 to $\sin^{-1}x = \int_0^x \frac{dx}{\sqrt{1-x^2}}$, etc., to find developments of

(α) $\sin^{-1}x$, $\quad(\beta)$ $\tan^{-1}x$, $\quad(\gamma)$ $\sinh^{-1}x$,

(δ) $\log\frac{1+x}{1-x}$, $\quad(\epsilon)$ $\int_0^x e^{-x^2}dx$, $\quad(\zeta)$ $\int_0^x \frac{\sin x}{x}dx$.

In all these cases the results may be found if desired to n terms.

8. Show that the remainder in the Maclaurin development of e^x is less than $x^n e^x/n!$; and hence that the error introduced by disregarding the remainder in computing e^x is less than $x^n e^x/n!$. How many terms will suffice to compute e to four decimals? How many for e^5 and for $e^{0.1}$?

9. Show that the error introduced by disregarding the remainder in computing $\log(1+x)$ is not greater than x^n/n if $x > 0$. How many terms are required for the computation of $\log 1\frac{1}{2}$ to four places? of $\log 1.2$? Compute the latter.

10. The hypotenuse of a triangle is 20 and one angle is $31°$. Find the sides by expanding $\sin x$ and $\cos x$ about $a = \frac{1}{6}\pi$ as linear functions of $x - \frac{1}{6}\pi$. Examine the term in $(x - \frac{1}{6}\pi)^2$ to find a maximum value to the error introduced by neglecting it.

11. Compute to 6 places: (α) $e^{\frac{1}{3}}$, (β) $\log 1.1$, (γ) $\sin 30'$, (δ) $\cos 30'$. During the computation one place more than the desired number should be carried along in the arithmetic work for safety.

12. Show that the remainder for $\log(1+x)$ is less than $x^n/n(1+x)^n$ if $x < 0$. Compute (α) $\log 0.9$ to 5 places, (β) $\log 0.8$ to 4 places.

13. Show that the remainder for $\tan^{-1}x$ is less than x^n/n where n may always be taken as odd. Compute to 4 places $\tan^{-1}\frac{1}{5}$.

14. The relation $\frac{1}{4}\pi = \tan^{-1}1 = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}$, enables $\frac{1}{4}\pi$ to be found easily from the series for $\tan^{-1}x$. Find $\frac{1}{4}\pi$ to 7 places (intermediate work carried to 8 places).

15. *Computation of logarithms.* (α) If $a = \log\frac{10}{9}$, $b = \log\frac{25}{24}$, $c = \log\frac{81}{80}$, then

$$\log 2 = 7a - 2b + 3c, \qquad \log 3 = 11a - 3b + 5c, \qquad \log 5 = 16a - 4b + 7c.$$

Now $a = -\log(1-\frac{1}{10})$, $b = -\log(1-\frac{4}{100})$, $c = \log(1+\frac{1}{80})$ are readily computed and hence $\log 2$, $\log 3$, $\log 5$ may be found. Carry the calculations of a, b, c to 10 places and deduce the logarithms of 2, 3, 5, 10, retaining only 8 places. Compare Peirce's "Tables," p. 109.

(β) Show that the error in the series for $\log\dfrac{1+x}{1-x}$ is less than $\dfrac{2x^n}{n(1-x)^n}$. Compute $\log 2$ corresponding to $x=\frac{1}{3}$ to 4 places, $\log 1\frac{2}{3}$ to 5 places, $\log 1\frac{2}{5}$ to 6 places.

(γ) Show $\log\dfrac{p}{q} = 2\left[\dfrac{p-q}{p+q} + \dfrac{1}{3}\left(\dfrac{p-q}{p+q}\right)^3 + \cdots + \dfrac{1}{2n-1}\left(\dfrac{p-q}{p+q}\right)^{2n-1} + R_{2n+1}\right]$, give an estimate of R_{2n+1}, and compute to 10 figures $\log 3$ and $\log 7$ from $\log 2$ and $\log 5$ of Peirce's "Tables" and from

$$4\log 3 - 4\log 2 - \log 5 = \log\frac{81}{80}, \qquad 4\log 7 - 5\log 2 - \log 3 - 2\log 5 = \log\frac{7^4}{7^4-1}.$$

16. Compute Ex. 7 (ϵ) to 4 places for $x=1$ and to 6 places for $x=\frac{1}{2}$.

17. Compute $\sin^{-1}0.1$ to seconds and $\sin^{-1}\frac{1}{3}$ to minutes.

18. Show that in the expansion of $(1+x)^k$ the remainder, as x is $>$ or <0, is

$$R_n < \left|\frac{k\cdot(k-1)\cdots(k-n+1)}{1\cdot 2\cdots n}x^n\right| \quad\text{or}\quad R_n < \left|\frac{k\cdot(k-1)\cdots(k-n+1)}{1\cdot 2\cdots n}\frac{x^n}{(1+x)^{n-k}}\right|, \; n>k.$$

Hence compute to 5 figures $\sqrt{103}$, $\sqrt{98}$, $\sqrt[3]{28}$, $\sqrt[5]{250}$, $\sqrt[10]{1000}$.

19. Sometimes the remainder cannot be readily found but the terms of the expansion appear to be diminishing so rapidly that all after a certain point appear negligible. Thus use Peirce's "Tables," Nos. 774–789, to compute to four places (estimated) the values of $\tan 6°$, $\log\cos 10°$, $\csc 3°$, $\sec 2°$.

20. Find to within 1% the area under $\cos(x^2)$ and $\sin(x^2)$ from 0 to $\frac{1}{2}\pi$.

21. A unit magnetic pole is placed at a distance L from the center of a magnet of pole strength M and length $2l$, where l/L is small. Find the force on the pole if (α) the pole is in the line of the magnet and if (β) it is in the perpendicular bisector.

Ans. (α) $\dfrac{4Ml}{L^3}(1+\epsilon)$ with ϵ about $2\left(\dfrac{l}{L}\right)^2$, ($\beta$) $\dfrac{2Ml}{L^3}(1-\epsilon)$ with ϵ about $\dfrac{3}{2}\left(\dfrac{l}{L}\right)^2$.

22. The formula for the distance of the horizon is $D=\sqrt{\frac{3}{2}h}$ where D is the distance in miles and h is the altitude of the observer in feet. Prove the formula and show that the error is about $\frac{1}{2}$% for heights up to a few miles. Take the radius of the earth as 3960 miles.

23. Find an approximate formula for the dip of the horizon in minutes below the horizontal if h in feet is the height of the observer.

24. If S is a circular arc and C its chord and c the chord of half the arc, prove $S=\frac{1}{3}(8c-C)(1+\epsilon)$ where ϵ is about $S^4/7680R^4$ if R is the radius.

25. If two quantities differ from each other by a small fraction ϵ of their value, show that their geometric mean will differ from their arithmetic mean by about $\frac{1}{8}\epsilon^2$ of its value.

26. The algebraic method may be applied to finding expansions of some functions which become infinite. (Thus if the series for $\cos x$ and $\sin x$ be divided to find $\cot x$, the initial term is $1/x$ and becomes infinite at $x=0$ just as $\cot x$ does.

Such expansions are not Maclaurin developments but are analogous to them. The function $x \cot x$ would, however, have a Maclaurin development and the expansion found for $\cot x$ is this development divided by x.) Find the developments about $x = 0$ to terms in x^4 for

(α) $\cot x$, (β) $\cot^2 x$, (γ) $\csc x$, (δ) $\csc^3 x$,
(ϵ) $\cot x \csc x$, (ζ) $1/(\tan^{-1} x)^2$, (η) $(\sin x - \tan x)^{-1}$

27. Obtain the expansions:

(α) $\log \sin x = \log x - \frac{1}{6} x^2 - \frac{1}{180} x^4 + R$, (β) $\log \tan x = \log x + \frac{1}{3} x^2 + \frac{7}{90} x^4 + \cdots$,
(γ) likewise for $\log \operatorname{vers} x$.

33. Indeterminate forms, infinitesimals, infinites. If two functions $f(x)$ and $\phi(x)$ are defined for $x = a$ and if $\phi(a) \neq 0$, the quotient f/ϕ is defined for $x = a$. But if $\phi(a) = 0$, the quotient f/ϕ is not defined for a. If in this case f and ϕ are defined and continuous in the neighborhood of a and $f(a) \neq 0$, the quotient will become infinite as $x \doteq a$; whereas if $f(a) = 0$, the behavior of the quotient f/ϕ is not immediately apparent but gives rise to the indeterminate form $0/0$. In like manner if f and ϕ become infinite at a, the quotient f/ϕ is not defined, as neither its numerator nor its denominator is defined; thus arises the indeterminate form ∞/∞. The question of determining or evaluating an indeterminate form is merely the question of finding out whether the quotient f/ϕ approaches a limit (and if so, what limit) or becomes positively or negatively infinite when x approaches a.

THEOREM. *L'Hospital's Rule.* If the functions $f(x)$ and $\phi(x)$, which give rise to the indeterminate form $0/0$ or ∞/∞ when $x \doteq a$, are continuous and differentiable in the interval $a < x \leqq b$ and if b can be taken so near to a that $\phi'(x)$ does not vanish in the interval and if the quotient f'/ϕ' of the derivatives approaches a limit or becomes positively or negatively infinite as $x \doteq a$, then the quotient f/ϕ will approach that limit or become positively or negatively infinite as the case may be. Hence *an indeterminate form $0/0$ or ∞/∞ may be replaced by the quotient of the derivatives of numerator and denominator.*

CASE I. $f(a) = \phi(a) = 0$. The proof follows from Cauchy's Formula, Ex. 6, p. 49.

For
$$\frac{f(x)}{\phi(x)} = \frac{f(x) - f(a)}{\phi(x) - \phi(a)} = \frac{f'(\xi)}{\phi'(\xi)}, \qquad a < \xi < x.$$

Now if $x \doteq a$, so must ξ, which lies between x and a. Hence if the quotient on the right approaches a limit or becomes positively or negatively infinite, the same is true of that on the left. The necessity of inserting the restrictions that f and ϕ shall be continuous and differentiable and that ϕ' shall not have a root indefinitely near to a is apparent from the fact that Cauchy's Formula is proved only for functions that satisfy these conditions. If the derived form f'/ϕ' should also be indeterminate, the rule could again be applied and the quotient f''/ϕ'' would replace f'/ϕ' with the understanding that proper restrictions were satisfied by f', ϕ', and ϕ''.

CASE II. $f(a) = \phi(a) = \infty$. Apply Cauchy's Formula as follows:

$$\frac{f(x) - f(b)}{\phi(x) - \phi(b)} = \frac{f(x)}{\phi(x)} \frac{1 - f(b)/f(x)}{1 - \phi(b)/\phi(x)} = \frac{f'(\xi)}{\phi'(\xi)}, \qquad \begin{array}{l} a < x < b, \\ x < \xi < b, \end{array}$$

where the middle expression is merely a different way of writing the first. Now suppose that $f'(x)/\phi'(x)$ approaches a limit when $x \doteq a$. It must then be possible to take b so near to a that $f'(\xi)/\phi'(\xi)$ differs from that limit by as little as desired, no matter what value ξ may have between a and b. Now as f and ϕ become infinite when $x \doteq a$, it is possible to take x so near to a that $f(b)/f(x)$ and $\phi(b)/\phi(x)$ are as near zero as desired. The second equation above then shows that $f(x)/\phi(x)$, multiplied by a quantity which differs from 1 by as little as desired, is equal to a quantity $f'(\xi)/\phi'(\xi)$ which differs from the limit of $f'(x)/\phi'(x)$ as $x \doteq a$ by as little as desired. Hence f/ϕ must approach the same limit as f'/ϕ'. Similar reasoning would apply to the supposition that f'/ϕ' became positively or negatively infinite, and the theorem is proved. It may be noted that, by Theorem 16 of § 27, the form f'/ϕ' is sure to be indeterminate. The advantage of being able to differentiate therefore lies wholly in the possibility that the new form be more amenable to algebraic transformation than the old.

The other indeterminate forms $0 \cdot \infty$, 0^0, 1^∞, ∞^0, $\infty - \infty$ may be reduced to the foregoing by various devices which may be indicated as follows:

$$0 \cdot \infty = \frac{0}{\frac{1}{\infty}} = \frac{\infty}{\frac{1}{0}}, \quad 0^0 = e^{\log 0^0} = e^{0 \log 0} = e^{0 \cdot \infty}, \quad \cdots, \quad \infty - \infty = \log e^\infty - \infty = \log \frac{e^\infty}{e^\infty}.$$

The case where the variable becomes infinite instead of approaching a finite value a is covered in Ex. 1 below. The theory is therefore completed.

Two methods which frequently may be used to shorten the work of evaluating an indeterminate form are *the method of E-functions* and *the application of Taylor's Formula*. By definition *an E-function for the point $x = a$ is any continuous function which approaches a finite limit other than 0 when $x \doteq a$.* Suppose then that $f(x)$ or $\phi(x)$ or both may be written as the products $E_1 f_1$ and $E_2 \phi_1$. Then the method of treating indeterminate forms need be applied only to f_1/ϕ_1 and the result multiplied by $\lim E_1/E_2$. For example,

$$\lim_{x \doteq a} \frac{x^3 - a^3}{\sin(x - a)} = \lim_{x \doteq a} (x^2 + ax + a^2) \lim_{x \doteq a} \frac{x - a}{\sin(x - a)} = 3a^2 \lim_{x \doteq a} \frac{x - a}{\sin(x - a)} = 3a^2.$$

Again, suppose that in the form 0/0 both numerator and denominator may be developed about $x = a$ by Taylor's Formula. The evaluation is immediate. Thus

$$\frac{\tan x - \sin x}{x^2 \log(1 + x)} = \frac{(x + \frac{1}{3}x^3 + Px^5) - (x - \frac{1}{6}x^3 + Qx^5)}{x^2(x - \frac{1}{2}x^2 + Rx^3)} = \frac{\frac{1}{2} + (P - Q)x^2}{1 - \frac{1}{2}x + Rx^2};$$

and now if $x \doteq 0$, the limit is at once shown to be simply $\frac{1}{2}$.

When the functions become infinite at $x = a$, the conditions requisite for Taylor's Formula are not present and there is no Taylor expansion. Nevertheless an expansion may sometimes be obtained by the algebraic method (§ 32) and may frequently be used to advantage. To illustrate, let it be required to evaluate $\cot x - 1/x$ which is of the form $\infty - \infty$ when $x \doteq 0$. Here

$$\cot x = \frac{\cos x}{\sin x} = \frac{1 + \frac{1}{2}x^2 + Px^4}{x - \frac{1}{6}x^3 + Qx^5} = \frac{1}{x} \frac{1 - \frac{1}{2}x^2 + Px^4}{1 - \frac{1}{6}x^2 + Qx^4} = \frac{1}{x}\left(1 - \frac{1}{3}x^2 + Sx^4\right),$$

where S remains finite when $x \doteq 0$. If this value be substituted for $\cot x$, then

$$\lim_{x \doteq 0}\left(\cot x - \frac{1}{x}\right) = \lim_{x \doteq 0}\left(\frac{1}{x} - \frac{1}{3}x + Sx^3 - \frac{1}{x}\right) = \lim_{x \doteq 0}\left(-\frac{1}{3}x + Sx^3\right) = 0.$$

34. *An infinitesimal is a variable which is ultimately to approach the limit zero; an infinite is a variable which is to become either positively or negatively infinite.* Thus the increments Δy and Δx are finite quantities, but when they are to serve in the definition of a derivative they must ultimately approach zero and hence may be called infinitesimals. The form $0/0$ represents the quotient of two infinitesimals;* the form ∞/∞, the quotient of two infinites; and $0 \cdot \infty$, the product of an infinitesimal by an infinite. If any infinitesimal α is chosen as the *primary infinitesimal*, a second infinitesimal β is said to be *of the same order* as α if the limit of the quotient β/α exists and is not zero when $\alpha \doteq 0$; whereas if the quotient β/α becomes zero, β is said to be an infinitesimal *of higher order* than α, but *of lower order* if the quotient becomes infinite. If in particular the limit β/α^n exists and is not zero when $\alpha \doteq 0$, then β is said to be *of the nth order relative to* α. The determination of the order of one infinitesimal relative to another is therefore essentially a problem in indeterminate forms. Similar definitions may be given in regard to infinites.

THEOREM. If the quotient β/α of two infinitesimals approaches a limit or becomes infinite when $\alpha \doteq 0$, the quotient β'/α' of two infinitesimals which differ respectively from β and α by infinitesimals of higher order will approach the same limit or become infinite.

THEOREM. *Duhamel's Theorem.* If the sum $\Sigma\alpha_i = \alpha_1 + \alpha_2 + \cdots + \alpha_n$ of n positive infinitesimals approaches a limit when their number n becomes infinite, the sum $\Sigma\beta_i = \beta_1 + \beta_2 + \cdots + \beta_n$, where each β_i differs uniformly from the corresponding α_i by an infinitesimal of higher order, will approach the same limit.

As $\alpha' - \alpha$ is of higher order than α and $\beta' - \beta$ of higher order than β,

$$\lim \frac{\alpha' - \alpha}{\alpha} = 0, \quad \lim \frac{\beta' - \beta}{\beta} = 0 \quad \text{or} \quad \frac{\alpha'}{\alpha} = 1 + \eta, \quad \frac{\beta'}{\beta} = 1 + \zeta,$$

where η and ζ are infinitesimals. Now $\alpha' = \alpha(1 + \eta)$ and $\beta' = \beta(1 + \zeta)$. Hence

$$\frac{\beta'}{\alpha'} = \frac{\beta}{\alpha}\,\frac{1 + \zeta}{1 + \eta} \quad \text{and} \quad \lim \frac{\beta'}{\alpha'} = \lim \frac{\beta}{\alpha},$$

provided β/α approaches a limit; whereas if β/α becomes infinite, so will β'/α'. In a more complex fraction such as $(\beta - \gamma)/\alpha$ it is *not* permissible to replace β

* It cannot be emphasized too strongly that in the symbol $0/0$ the 0's are merely symbolic for a mode of variation just as ∞ is; they are not actual 0's and some other notation would be far preferable, likewise for $0 \cdot \infty$, 0^0, etc.

and γ individually by infinitesimals of higher order; for $\beta - \gamma$ may itself be of higher order than β or γ. Thus $\tan x - \sin x$ is an infinitesimal of the third order relative to x although $\tan x$ and $\sin x$ are only of the first order. To replace $\tan x$ and $\sin x$ by infinitesimals which differ from them by those of the second order or even of the third order would generally alter the limit of the ratio of $\tan x - \sin x$ to x^3 when $x \doteq 0$.

To prove Duhamel's Theorem the β's may be written in the form

$$\beta_i = \alpha_i(1 + \eta_i), \qquad i = 1, 2, \cdots, n, \qquad |\eta_i| < \epsilon,$$

where the η's are infinitesimals and where all the η's simultaneously may be made less than the assigned ϵ owing to the uniformity required in the theorem. Then

$$|(\beta_1 + \beta_2 + \cdots + \beta_n) - (\alpha_1 + \alpha_2 + \cdots + \alpha_n)| = |\eta_1\alpha_1 + \eta_2\alpha_2 + \cdots + \eta_n\alpha_n| < \epsilon\Sigma\alpha.$$

Hence the sum of the β's may be made to differ from the sum of the α's by less than $\epsilon\Sigma\alpha$, a quantity as small as desired, and as $\Sigma\alpha$ approaches a limit by hypothesis, so $\Sigma\beta$ must approach the same limit. The theorem may clearly be extended to the case where the α's are not all positive provided the sum $\Sigma|\alpha_i|$ of the absolute values of the α's approaches a limit.

35. If $y = f(x)$, the *differential* of y is defined as

$$dy = f'(x)\,\Delta x, \qquad \text{and hence} \qquad dx = 1 \cdot \Delta x. \tag{4}$$

From this definition of dy and dx it appears that $dy/dx = f'(x)$, where the quotient dy/dx is the quotient of two finite quantities of which dx may be assigned at pleasure. This is true if x is the independent variable. If x and y are both expressed in terms of t,

$$x = x(t), \qquad y = y(t), \qquad dx = D_t x\,dt, \qquad dy = D_t y\,dt;$$

and
$$\frac{dy}{dx} = \frac{D_t y}{D_t x} = D_x y, \qquad \text{by virtue of (1), § 2.}$$

From this appears the important theorem: *The quotient dy/dx is the derivative of y with respect to x no matter what the independent variable may be.* It is this theorem which really justifies writing the derivative as a fraction and treating the component differentials according to the rules of ordinary fractions. For higher derivatives this is not so, as may be seen by reference to Ex. 10.

As Δy and Δx are regarded as infinitesimals in defining the derivative, it is natural to regard dy and dx as infinitesimals. The difference $\Delta y - dy$ may be put in the form

$$\Delta y - dy = \left[\frac{f(x + \Delta x) - f(x)}{\Delta x} - f'(x)\right]\Delta x, \tag{5}$$

wherein it appears that, when $\Delta x \doteq 0$, the bracket approaches zero. Hence arises the theorem: *If x is the independent variable and if Δy and dy are regarded as infinitesimals, the difference $\Delta y - dy$ is an infinitesimal of higher order than Δx.* This has an application to the

subject of change of variable in a definite integral. For if $x = \phi(t)$, then $dx = \phi'(t)\,dt$, and apparently

$$\int_a^b f(x)\,dx = \int_{t_1}^{t_2} f[\phi(t)]\,\phi'(t)\,dt,$$

where $\phi(t_1) = a$ and $\phi(t_2) = b$, so that t ranges from t_1 to t_2 when x ranges from a to b.

But this substitution is too hasty; for the dx written in the integrand is really Δx, which differs from dx by an infinitesimal of higher order when x is not the independent variable. The true condition may be seen by comparing the two sums

$$\sum f(x_i)\,\Delta x_i, \qquad \sum f[\phi(t_i)]\,\phi'(t_i)\,\Delta t_i, \qquad \Delta t = dt,$$

the limits of which are the two integrals above. Now as Δx differs from $dx = \phi'(t)\,dt$ by an infinitesimal of higher order, so $f(x)\,\Delta x$ will differ from $f[\phi(t)]\,\phi'(t)\,dt$ by an infinitesimal of higher order, and with the proper assumptions as to continuity the difference will be uniform. Hence if the infinitesimals $f(x)\,\Delta x$ be all positive, Duhamel's Theorem may be applied to justify the formula for change of variable. To avoid the restriction to positive infinitesimals it is well to replace Duhamel's Theorem by the new

Theorem. *Osgood's Theorem.* Let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be n infinitesimals and let α_i differ uniformly by infinitesimals of higher order than Δx from the elements $f(x_i)\,\Delta x_i$ of the integrand of a definite integral $\int_a^b f(x)\,dx$, where f is continuous; then the sum $\Sigma\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_n$ approaches the value of the definite integral as a limit when the number n becomes infinite.

Let $\alpha_i = f(x_i)\,\Delta x_i + \zeta_i \Delta x_i$, where $|\zeta_i| < \epsilon$ owing to the uniformity demanded.

Then $$\left|\sum \alpha_i - \sum f(x_i)\,\Delta x_i\right| = \left|\sum \zeta_i \Delta x_i\right| < \epsilon \sum \Delta x_i = \epsilon(b-a).$$

But as f is continuous, the definite integral exists and one can make

$$\left|\sum f(x_i)\,\Delta x_i - \int_a^b f(x)\,dx\right| < \epsilon, \quad \text{and hence} \quad \left|\sum \alpha_i - \int_a^b f(x)\,dx\right| < \epsilon(b-a+1).$$

It therefore appears that $\Sigma\alpha_i$ may be made to differ from the integral by as little as desired, and $\Sigma\alpha_i$ must then approach the integral as a limit. Now if this theorem be applied to the case of the change of variable and if it be assumed that $f[\phi(t)]$ and $\phi'(t)$ are continuous, the infinitesimals Δx_i and $dx_i = \phi'(t_i)\,dt_i$ will differ uniformly (compare Theorem 18 of § 27 and the above theorem on $\Delta y - dy$) by an infinitesimal of higher order, and so will the infinitesimals $f(x_i)\,\Delta x_i$ and $f[\phi(t_i)]\,\phi'(t_i)\,dt_i$. Hence the change of variable suggested by the hasty substitution is justified.

EXERCISES

1. Show that l'Hospital's Rule applies to evaluating the indeterminate form $f(x)/\phi(x)$ when x becomes infinite and both f and ϕ either become zero or infinite.

2. Evaluate the following forms by differentiation. Examine the quotients for left-hand and for right-hand approach; sketch the graphs in the neighborhood of the points.

(α) $\lim\limits_{x \doteq 0} \dfrac{a^x - b^x}{x}$, (β) $\lim\limits_{x \doteq \frac{1}{4}\pi} \dfrac{\tan x - 1}{x - \frac{1}{4}\pi}$, (γ) $\lim\limits_{x \doteq 0} x \log x$,

(δ) $\lim\limits_{x \doteq \infty} xe^{-x}$, (ϵ) $\lim\limits_{x \doteq 0} (\cot x)^{\sin x}$, (ζ) $\lim\limits_{x \doteq 1} x^{\frac{1}{1-x}}$.

3. Evaluate the following forms by the method of expansions:

(α) $\lim\limits_{x \doteq 0} \left(\dfrac{1}{x^2} - \cot^2 x\right)$, (β) $\lim\limits_{x \doteq 0} \dfrac{e^x - e^{\tan x}}{x - \tan x}$, (γ) $\lim\limits_{x \doteq 1} \dfrac{\log x}{1 - x}$,

(δ) $\lim\limits_{x \doteq 0} (\operatorname{csch} x - \csc x)$, (ϵ) $\lim\limits_{x \doteq 0} \dfrac{x \sin(\sin x) - \sin^2 x}{x^6}$, (ζ) $\lim\limits_{x \doteq 0} \dfrac{e^x - e^{-x} - 2x}{x - \sin x}$.

4. Evaluate by any method:

(α) $\lim\limits_{x \doteq 0} \dfrac{e^x - e^{-x} + 2 \sin x - 4x}{x^5}$, (β) $\lim\limits_{x \doteq 0} \left(\dfrac{\tan x}{x}\right)^{\frac{1}{x^2}}$,

(γ) $\lim\limits_{x \doteq 0} \dfrac{x \cos^3 x - \log(1 + x) - \sin^{-1} \frac{1}{2} x^2}{x^3}$, (δ) $\lim\limits_{x \doteq \frac{1}{2}\pi} \dfrac{\log(x - \frac{1}{2}\pi)}{\tan x}$,

$$(\epsilon)\ \lim_{x = \infty} \left[x\left(1 + \frac{1}{x}\right)^x - ex^2 \log\left(1 + \frac{1}{x}\right)\right].$$

5. Give definitions for order as applied to infinites, noting that higher order would mean becoming infinite to a greater degree just as it means becoming zero to a greater degree for infinitesimals. State and prove the theorem relative to quotients of infinites analogous to that given in the text for infinitesimals. State and prove an analogous theorem for the product of an infinitesimal and infinite.

6. Note that if the quotient of two infinites has the limit 1, the difference of the infinites is an infinite of lower order. Apply this to the proof of the resolution in partial fractions of the quotient $f(x)/F(x)$ of two polynomials in case the roots of the denominator are all real. For if $F(x) = (x - a)^k F_1(x)$, the quotient is an infinite of order k in the neighborhood of $x = a$; but the difference of the quotient and $f(a)/(x - a)^k F_1(a)$ will be of lower integral order — and so on.

7. Show that when $x = +\infty$, the function e^x is an infinite of higher order than x^n no matter how large n. Hence show that if $P(x)$ is any polynomial, $\lim\limits_{x \doteq \infty} P(x)e^{-x} = 0$ when $x = +\infty$.

8. Show that $(\log x)^m$ when x is infinite is a weaker infinite than x^n no matter how large m or how small n, supposed positive, may be. What is the graphical interpretation?

9. If P is a polynomial, show that $\lim\limits_{x \doteq 0} P\left(\dfrac{1}{x}\right)e^{-\frac{1}{x^2}} = 0$. Hence show that the Maclaurin development of $e^{-\frac{1}{x^2}}$ is $f(x) = e^{-\frac{1}{x^2}} = \dfrac{x^n}{n!} f^{(n)}(\theta x)$ if $f(0)$ is defined as 0.

10. The higher differentials are defined as $d^ny = f^{(n)}(x)\,(dx)^n$ where x is taken as the independent variable. Show that $d^kx = 0$ for $k > 1$ if x is the independent variable. Show that the higher derivatives D_x^2y, $D_x^3y, \cdots$ are not the quotients d^2y/dx^2, $d^3y/dx^3, \cdots$ if x and y are expressed in terms of a third variable, but that the relations are

$$D_x^2y = \frac{d^2y dx - d^2x dy}{dx^3}, \qquad D_x^3y = \frac{dx\,(dx d^3y - dy d^3x) - 3\,d^2x\,(dx d^2y - dy d^2x)}{dx^5}, \quad \cdots.$$

The fact that the quotient d^ny/dx^n, $n > 1$, is not the derivative when x and y are expressed parametrically militates against the usefulness of the higher differentials and emphasizes the advantage of working with derivatives. The notation d^ny/dx^n is, however, used for the derivative. Nevertheless, as indicated in Exs. 16–19, higher differentials may be used if proper care is exercised.

11. Compare the conception of higher differentials with the work of Ex. 5, p. 48.

12. Show that in a circle the difference between an infinitesimal arc and its chord is of the third order relative to either arc or chord.

13. Show that if β is of the nth order with respect to α, and γ is of the first order with respect to α, then β is of the nth order with respect to γ.

14. Show that the order of a product of infinitesimals is equal to the sum of the orders of the infinitesimals when all are referred to the same primary infinitesimal α. Infer that in a product each infinitesimal may be replaced by one which differs from it by an infinitesimal of higher order than it without affecting the order of the product.

15. Let A and B be two points of a unit circle and let the angle AOB subtended at the center be the primary infinitesimal. Let the tangents at A and B meet at T, and OT cut the chord AB in M and the arc AB in C. Find the trigonometric expression for the infinitesimal difference $TC - CM$ and determine its order.

16. Compute $d^2(x \sin x) = (2\cos x - x\sin x)\,dx^2 + (\sin x + x\cos x)\,d^2x$ by taking the differential of the differential. Thus find the second derivative of $x \sin x$ if x is the independent variable and the second derivative with respect to t if $x = 1 + t^2$.

17. Compute the first, second, and third differentials, $d^2x \neq 0$.

(α) $x^2 \cos x$, (β) $\sqrt{1-x}\,\log(1-x)$, (γ) $xe^{2x}\sin x$.

18. In Ex. 10 take y as the independent variable and hence express D_x^2y, D_x^3y in terms of D_yx, D_y^2x. Cf. Ex. 10, p. 14.

19. Make the changes of variable in Exs. 8, 9, 12, p. 14, by the method of differentials, that is, by replacing the derivatives by the corresponding differential expressions where x is not assumed as independent variable and by replacing these differentials by their values in terms of the new variables where the higher differentials of the new independent variable are set equal to 0.

20. Reconsider some of the exercises at the end of Chap. I, say, 17–19, 22, 23, 27, from the point of view of Osgood's Theorem instead of the Theorem of the Mean.

21. Find the areas of the bounding surfaces of the solids of Ex. 11, p. 18.

22. Assume the law $F = kmm'/r^2$ of attraction between particles. Find the attraction of :

(α) a circular wire of radius a and of mass M on a particle m at a distance r from the center of the wire along a perpendicular to its plane ; *Ans.* $kMmr(a^2 + r^2)^{-\frac{3}{2}}$.

(β) a circular disk, etc., as in (α) ; *Ans.* $2kMma^{-2}(1 - r/\sqrt{r^2 + a^2})$.

(γ) a semicircular wire on a particle at its center ; *Ans.* $2kMm/\pi a^2$.

(δ) a finite rod upon a particle not in the line of the rod. The answer should be expressed in terms of the angle the rod subtends at the particle.

(ϵ) two parallel equal rods, forming the opposite sides of a rectangle, on each other.

23. Compare the method of derivatives (§ 7), the method of the Theorem of the Mean (§ 17), and the method of infinitesimals above as applied to obtaining the formulas for (α) area in polar coördinates, (β) mass of a rod of variable density, (γ) pressure on a vertical submerged bulkhead, (δ) attraction of a rod on a particle. Obtain the results by each method and state which method seems preferable for each case.

24. Is the substitution $dx = \phi'(t)\,dt$ in the indefinite integral $\int f(x)\,dx$ to obtain the indefinite integral $\int f[\phi(t)]\,\phi'(t)\,dt$ justifiable immediately ?

36. Infinitesimal analysis. To work rapidly in the applications of calculus to problems in geometry and physics and to follow readily the books written on those subjects, it is necessary to have some familiarity with working directly with infinitesimals. It is possible by making use of the Theorem of the Mean and allied theorems to retain in every expression its complete exact value ; but if that expression is an infinitesimal which is ultimately to enter into a quotient or a limit of a sum, any infinitesimal which is of higher order than that which is ultimately kept will not influence the result and may be discarded at any stage of the work if the work may thereby be simplified. A few theorems worked through by the infinitesimal method will serve partly to show how the method is used and partly to establish results which may be of use in further work. The theorems which will be chosen are :

1. The increment Δx and the differential dx of a variable differ by an infinitesimal of higher order than either.

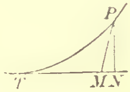
2. If a tangent is drawn to a curve, the perpendicular from the curve to the tangent is of higher order than the distance from the foot of the perpendicular to the point of tangency.

3. An infinitesimal arc differs from its chord by an infinitesimal of higher order relative to the arc.

4. If one angle of a triangle, none of whose angles are infinitesimal, differs infinitesimally from a right angle and if h is the side opposite and if ϕ is another angle of the triangle, then the side opposite ϕ is $h \sin \phi$ except for an infinitesimal of the second order and the adjacent side is $h \cos \phi$ except for an infinitesimal of the first order.

The first of these theorems has been proved in § 35. The second follows from it and from the idea of tangency. For take the x-axis coincident with the tangent or parallel to it. Then the perpendicular is Δy and the distance from its foot to the point of tangency is Δx. The quotient $\Delta y/\Delta x$ approaches 0 as its limit because the tangent is horizontal; and the theorem is proved. *The theorem would remain true if the perpendicular were replaced by a line making a constant angle with the tangent and the distance from the point of tangency to the foot of the perpendicular were replaced by the distance to the foot of the oblique line.* For if $\angle PMN = \theta$,

$$\frac{PM}{TM} = \frac{PN \csc\theta}{TN - PN \cot\theta} = \frac{PN}{TN} \frac{\csc\theta}{1 - \dfrac{PN}{TN}\cot\theta},$$

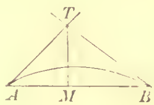


and therefore when P approaches T with θ constant, PM/TM approaches zero and PM is of higher order than TM.

The third theorem follows without difficulty from the assumption or theorem that the arc has a length intermediate between that of the chord and that of the sum of the two tangents at the ends of the chord. Let θ_1 and θ_2 be the angles between the chord and the tangents. Then

$$\frac{s - AB}{AM + MB} < \frac{AT + TB - AB}{AM + MB} = \frac{AM(\sec\theta_1 - 1) + MB(\sec\theta_2 - 1)}{AM + MB}. \qquad (6)$$

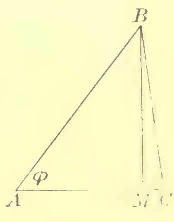
Now as AB approaches 0, both $\sec\theta_1 - 1$ and $\sec\theta_2 - 1$ approach 0 and their coefficients remain necessarily finite. Hence the difference between the arc and the chord is an infinitesimal of higher order than the chord. As the arc and chord are therefore of the same order, the difference is of higher order than the arc. This result enables one to replace the arc by its chord and vice versa in discussing infinitesimals of the first order, and for such purposes to consider an infinitesimal arc as straight. In discussing infinitesimals of the second order, this substitution would not be permissible except in view of the further theorem given below in § 37, and even then the substitution will hold only as far as the lengths of arcs are concerned and not in regard to directions.



For the fourth theorem let θ be the angle by which C departs from 90° and with the perpendicular BM as radius strike an arc cutting BC. Then by trigonometry

$$AC = AM + MC = h\cos\phi + BM\tan\theta,$$
$$BC = h\sin\phi + BM(\sec\theta - 1).$$

Now $\tan\theta$ is an infinitesimal of the first order with respect to θ; for its Maclaurin development begins with θ. And $\sec\theta - 1$ is an infinitesimal of the second order; for its development begins with a term in θ^2. The theorem is therefore proved. This theorem is frequently applied to infinitesimal triangles, that is, triangles in which h is to approach 0.



37. As a further discussion of the third theorem it may be recalled that by definition the length of the arc of a curve is the limit of the length of an inscribed polygon, namely,

$$s = \lim_{n=\infty} \left(\sqrt{\Delta x_1^2 + \Delta y_1^2} + \sqrt{\Delta x_2^2 + \Delta y_2^2} + \cdots + \sqrt{\Delta x_n^2 + \Delta y_n^2}\right).$$

Now $$\sqrt{\Delta x^2 + \Delta y^2} - \sqrt{dx^2 + dy^2} = \frac{\Delta x^2 + \Delta y^2 - dx^2 - dy^2}{\sqrt{\Delta x^2 + \Delta y^2} + \sqrt{dx^2 + dy^2}}$$
$$= \frac{(\Delta x - dx)(\Delta x + dx) + (\Delta y - dy)(\Delta y + dy)}{\sqrt{\Delta x^2 + \Delta y^2} + \sqrt{dx^2 + dy^2}},$$

and $$\frac{\sqrt{\Delta x^2 + \Delta y^2} - \sqrt{dx^2 + dy^2}}{\sqrt{\Delta x^2 + \Delta y^2}} = \frac{(\Delta x - dx)}{\sqrt{\Delta x^2 + \Delta y^2}} \frac{\Delta x + dx}{\sqrt{\Delta x^2 + \Delta y^2} + \sqrt{dx^2 + dy^2}}$$
$$+ \frac{(\Delta y - dy)}{\sqrt{\Delta x^2 + \Delta y^2}} \frac{\Delta y + dy}{\sqrt{\Delta x^2 + \Delta y^2} + \sqrt{dx^2 + dy^2}}.$$

But $\Delta x - dx$ and $\Delta y - dy$ are infinitesimals of higher order than Δx and Δy. Hence the right-hand side must approach zero as its limit and hence $\sqrt{\Delta x^2 + \Delta y^2}$ differs from $\sqrt{dx^2 + dy^2}$ by an infinitesimal of higher order and may replace it in the sum

$$s = \lim_{n=\infty} \sum \sqrt{\Delta x_i^2 + \Delta y_i^2} = \lim_{n=\infty} \sum \sqrt{dx^2 + dy^2} = \int_{x_0}^{x_1} \sqrt{1 + y'^2}\, dx.$$

The length of the arc measured from a fixed point to a variable point is a function of the upper limit and the differential of arc is

$$ds = d\int_{x_0}^{x} \sqrt{1 + y'^2}\, dx = \sqrt{1 + y'^2}\, dx = \sqrt{dx^2 + dy^2}.$$

To find the order of the difference between the arc and its chord let the origin be taken at the initial point and the x-axis tangent to the curve at that point. The expansion of the arc by Maclaurin's Formula gives

$$s(x) = s(0) + xs'(0) + \tfrac{1}{2}x^2 s''(0) + \tfrac{1}{6}x^3 s'''(\theta x),$$

where $s(0) = 0$, $\quad s'(0) = \sqrt{1 + y'^2}\,\big]_0 = 1$, $\quad s''(0) = \left.\dfrac{y'y''}{\sqrt{1 + y'^2}}\right]_0 = 0.$

Owing to the choice of axes, the expansion of the curve reduces to

$$y = f(x) = y(0) + xy'(0) + \tfrac{1}{2}x^2 y''(\theta x) = \tfrac{1}{2}x^2 y''(\theta x),$$

and hence the chord of the curve is

$$c(x) = \sqrt{x^2 + y^2} = x\sqrt{1 + \tfrac{1}{4}x^2[y''(\theta x)]^2} = x(1 + x^2 P),$$

where P is a complicated expression arising in the expansion of the radical by Maclaurin's Formula. The difference

$$s(x) - c(x) = [x + \tfrac{1}{6}x^3 s'''(\theta x)] - [x(1 + x^2 P)] = x^3(\tfrac{1}{6}s'''(\theta x) - P).$$

This is an infinitesimal of at least the third order relative to x. Now as both $s(x)$ and $c(x)$ are of the first order relative to x, it follows that the difference $s(x) - c(x)$ must also be of the third order relative to either $s(x)$ or $c(x)$. Note that the proof assumes that y'' is finite at the point considered. This result, which has been found analytically, follows more simply though perhaps less rigorously from the fact that $\sec\theta_1 - 1$ and $\sec\theta_2 - 1$ in (6) are infinitesimals of the second order with θ_1 and θ_2.

38. The theory of *contact of plane curves* may be treated by means of Taylor's Formula and stated in terms of infinitesimals. Let two curves $y = f(x)$ and $y = g(x)$ be tangent at a given point and let the

origin be chosen at that point with the x-axis tangent to the curves. The Maclaurin developments are

$$y = f(x) = \frac{1}{2}f''(0)x^2 + \cdots + \frac{1}{(n-1)!}x^{n-1}f^{(n-1)}(0) + \frac{1}{n!}x^{(n)}f^{(n)}(0) + \cdots$$

$$y = g(x) = \frac{1}{2}g''(0)x^2 + \cdots + \frac{1}{(n-1)!}x^{n-1}g^{(n-1)}(0) + \frac{1}{n!}x^n g^{(n)}(0) + \cdots.$$

If these developments agree up to but not including the term in x^n, the difference between the ordinates of the curves is

$$f(x) - g(x) = \frac{1}{n!}x^n[f^{(n)}(0) - g^{(n)}(0)] + \cdots, \qquad f^{(n)}(0) \neq g^{(n)}(0),$$

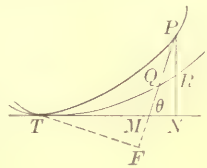
and is an infinitesimal of the nth order with respect to x. The curves are then said to have *contact of order* $n-1$ at their point of tangency. In general when two curves are tangent, the derivatives $f''(0)$ and $g''(0)$ are unequal and the curves have simple contact or *contact of the first order*.

The problem may be stated differently. Let PM be a line which makes a constant angle θ with the x-axis. Then, when P approaches T, if RQ be regarded as straight, the proportion

$$\lim (PR : PQ) = \lim (\sin \angle PQR : \sin \angle PRQ) = \sin\theta : 1$$

shows that PR and PQ are of the same order. Clearly also the lines TM and TN are of the same order. Hence if

$$\lim \frac{PR}{(TN)^n} \neq 0, \infty, \quad \text{then} \quad \lim \frac{PQ}{(TM)^n} \neq 0, \infty.$$



Hence if two curves have contact of the $(n-1)$st order, the segment of a line intercepted between the two curves is of the nth order with respect to the distance from the point of tangency to its foot. It would also be of the nth order with respect to the perpendicular TF from the point of tangency to the line.

In view of these results it is not necessary to assume that the two curves have a special relation to the axis. Let two curves $y = f(x)$ and $y = g(x)$ intersect when $x = a$, and assume that the tangents at that point are not parallel to the y-axis. Then

$$y = y_0 + (x-a)f'(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{(x-a)^n}{n!}f^{(n)}(a) + \cdots$$

$$y = y_0 + (x-a)g'(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{(x-a)^n}{n!}g^{(n)}(a) + \cdots$$

will be the Taylor developments of the two curves. If the difference of the ordinates for equal values of x is to be an infinitesimal of the nth order with respect to $x-a$ which is the perpendicular from the point of tangency to the ordinate, then the Taylor developments must agree up to but not including the terms in x^n. This is the condition for contact of order $n-1$.

As the difference between the ordinates is

$$f(x)-g(x)=\frac{1}{n!}(x-a)^n\left[f^{(n)}(a)-g^{(n)}(a)\right]+\cdots,$$

the difference will change sign or keep its sign when x passes through a according as n is odd or even, because for values sufficiently near to x the higher terms may be neglected. Hence *the curves will cross each other if the order of contact is even*, but *will not cross each other if the order of contact is odd.* If the values of the ordinates are equated to find the points of intersection of the two curves, the result is

$$0=\frac{1}{n!}(x-a)^n\left\{\left[f^{(n)}(a)-g^{(n)}(a)\right]+\cdots\right\}$$

and shows that $x=a$ is a root of multiplicity n. Hence it is said that two curves have in common as many coincident points as the order of their contact plus one. This fact is usually stated more graphically by saying that *the curves have n consecutive points in common.* It may be remarked that what Taylor's development carried to n terms does, is to give a polynomial which has contact of order $n-1$ with the function that is developed by it.

As a problem on contact consider the determination of the circle which shall have contact of the second order with a curve at a given point (a, y_0). Let

$$y=y_0+(x-a)f'(a)+\tfrac{1}{2}(x-a)^2f''(a)+\cdots$$

be the development of the curve and let $y'=f'(a)=\tan\tau$ be the slope. If the circle is to have contact with the curve, its center must be at some point of the normal. Then if R denotes the assumed radius, the equation of the circle may be written as

$$(x-a)^2+2R\sin\tau\,(x-a)+(y-y_0)^2-2R\cos\tau\,(y-y_0)=0,$$

where it remains to determine R so that the development of the circle will coincide with that of the curve as far as written. Differentiate the equation of the circle.

$$\frac{dy}{dx}=-\frac{R\sin\tau+(x-a)}{R\cos\tau-(y-y_0)},\qquad \left(\frac{dy}{dx}\right)_{a,\,y_0}=\tan\tau=f'(a).$$

$$\frac{d^2y}{dx^2}=-\frac{[R\cos\tau-(y-y_0)]^2+[R\sin\tau+(x-a)]^2}{[R\cos\tau-(y-y_0)]^3},\qquad \left(\frac{d^2y}{dx^2}\right)_{a,\,y_0}=-\frac{1}{R\cos^3\tau}.$$

and $$y=y_0+(x-a)f'(a)+\tfrac{1}{2}(x-a)^2\frac{-1}{R\cos^3\tau}+\cdots$$

is the development of the circle. The equation of the coefficients of $(x-a)^2$,

$$\frac{1}{R\cos^3\tau}=f''(a), \quad \text{gives} \quad R=\frac{\sec^3\tau}{f''(a)}=\frac{[1+[f'(a)]^2]^{\frac{3}{2}}}{f''(a)}.$$

This is the well known formula for the radius of curvature and shows that the circle of curvature has contact of at least the second order with the curve. The circle is sometimes called the osculating circle instead of the circle of curvature.

39. Three theorems, one in geometry and two in kinematics, will now be proved to illustrate the direct application of the infinitesimal methods to such problems. The choice will be:

1. The tangent to the ellipse is equally inclined to the focal radii drawn to the point of contact.

2. The displacement of any rigid body in a plane may be regarded at any instant as a rotation through an infinitesimal angle about some point unless the body is moving parallel to itself.

3. The motion of a rigid body in a plane may be regarded as the rolling of one curve upon another.

For the first problem consider a secant PP' which may be converted into a tangent TT' by letting the two points approach until they coincide. Draw the focal radii to P and P' and strike arcs with F and F' as centers. As $F'P+PF=F'P'+P'F=2a$, it follows that $NP=MP'$. Now consider the two triangles $PP'M$ and $P'PN$ nearly right-angled at M and N. The sides PP', PM, PN, $P'M$, $P'N$ are all infinitesimals of the same order and of the same order as the angles at F and F'. By proposition 4 of § 36

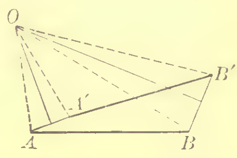
$$MP'=PP'\cos\angle PP'M+e_1, \qquad NP=PP'\cos\angle P'PN+e_2,$$

where e_1 and e_2 are infinitesimals relative to MP' and NP or PP'. Therefore

$$\lim[\cos\angle PP'M-\cos\angle P'PN]=\cos\angle TPF-\cos\angle T'PF'=\lim\frac{e_1-e_2}{PP'}=0,$$

and the two angles TPF' and $T'PF$ are proved to be equal as desired.

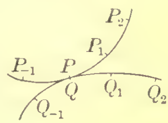
To prove the second theorem note first that if a body is rigid, its position is completely determined when the position AB of any rectilinear segment of the body is known. Let the points A and B of the body be describing curves AA' and BB' so that, in an infinitesimal interval of time, the line AB takes the neighboring position $A'B'$. Erect the perpendicular bisectors of the lines AA' and BB' and let them intersect at O. Then the triangles AOB and $A'OB'$ have the three sides of the one equal to the three sides of the other and are equal, and the second may be obtained from the first by a mere rotation about O through the angle $AOA'=BOB'$. Except for infinitesimals of higher order, the magnitude of the angle is AA'/OA or BB'/OB. Next let the interval of time approach 0 so that A' approaches A and B' approaches B. The perpendicular bisectors will approach



the normals to the arcs AA' and BB' at A and B, and the point O will approach the intersection of those normals.

The theorem may then be stated that: *At any instant of time the motion of a rigid body in a plane may be considered as a rotation through an infinitesimal angle about the intersection of the normals to the paths of any two of its points at that instant; the amount of the rotation will be the distance ds that any point moves divided by the distance of that point from the instantaneous center of rotation; the angular velocity about the instantaneous center will be this amount of rotation divided by the interval of time dt, that is, it will be v/r, where v is the velocity of any point of the body and r is its distance from the instantaneous center of rotation.* It is therefore seen that not only is the desired theorem proved, but numerous other details are found. As has been stated, the point about which the body is rotating at a given instant is called the *instantaneous center* for that instant.

As time goes on, the position of the instantaneous center will generally change. If at each instant of time the position of the center is marked on the moving plane or body, there results a locus which is called the *moving centrode* or *body centrode;* if at each instant the position of the center is also marked on a fixed plane over which the moving plane may be considered to glide, there results another locus which is called the *fixed centrode* or the *space centrode.* From these definitions it follows that at each instant of time the body centrode and the space centrode intersect at the instantaneous center for that instant. Consider a series of positions of the instantaneous center as $P_{-2}P_{-1}PP_1P_2$ marked in space and $Q_{-2}Q_{-1}QQ_1Q_2$ marked in the body. At a given instant two of the points, say P and Q, coincide; an instant later the body will have moved so as to bring Q_1 into coincidence with P_1; at an earlier instant Q_{-1} was coincident with P_{-1}. Now as the motion at the instant when P and Q are together is one of rotation through an infinitesimal angle about that point, the angle between PP_1 and QQ_1 is infinitesimal and the lengths PP_1 and QQ_1 are equal; for it is by the rotation about P and Q that Q_1 is to be brought into coincidence with P_1. Hence it follows 1° that the two centrodes are tangent and 2° that the distances $PP_1 = QQ_1$ which the point of contact moves along the two curves during an infinitesimal interval of time are the same, and this means that the two curves roll on one another without slipping — because the very idea of slipping implies that the point of contact of the two curves should move by different amounts along the two curves, the difference in the amounts being the amount of the slip. The third theorem is therefore proved.



EXERCISES

1. If a finite parallelogram is nearly rectangled, what is the order of infinitesimals neglected by taking the area as the product of the two sides? What if the figure were an isosceles trapezoid? What if it were any rectilinear quadrilateral all of whose angles differ from right angles by infinitesimals of the same order?

2. On a sphere of radius r the area of the zone between the parallels of latitude λ and $\lambda + d\lambda$ is taken as $2\pi r\cos\lambda \cdot r d\lambda$, the perimeter of the base times the slant height. Of what order relative to $d\lambda$ is the infinitesimal neglected? What if the perimeter of the middle latitude were taken so that $2\pi r^2\cos(\lambda + \frac{1}{2}d\lambda)d\lambda$ were assumed?

3. What is the order of the infinitesimal neglected in taking $4\pi r^2 dr$ as the volume of a hollow sphere of interior radius r and thickness dr? What if the mean radius were taken instead of the interior radius? Would any particular radius be best?

4. Discuss the length of a space curve $y = f(x)$, $z = g(x)$ analytically as the length of the plane curve was discussed in the text.

5. Discuss proposition 2, p. 68, by Maclaurin's Formula and in particular show that if the second derivative is continuous at the point of tangency, the infinitesimal in question is of the second order at least. How about the case of the tractrix

$$y = \frac{a}{2}\log\frac{a - \sqrt{a^2 - x^2}}{a + \sqrt{a^2 - x^2}} + \sqrt{a^2 - x^2},$$

and its tangent at the vertex $x = a$? How about $s(x) - c(x)$ of § 37?

6. Show that if two curves have contact of order $n-1$, their derivatives will have contact of order $n-2$. What is the order of contact of the kth derivatives $k < n - 1$?

7. State the conditions for maxima, minima, and points of inflection in the neighborhood of a point where $f^{(n)}(a)$ is the first derivative that does not vanish.

8. Determine the order of contact of these curves at their intersections:

$$(\alpha)\ \begin{array}{l} \sqrt{2}(x^2 + y^2 + 2) = 3(x + y) \\ 5x^2 - 6xy + 5y^2 = 8. \end{array} \qquad (\beta)\ \begin{array}{l} r^2 = a^2\cos 2\phi \\ y^2 = \frac{1}{3}a(a - x). \end{array} \qquad (\gamma)\ \begin{array}{l} x^2 + y^2 = y \\ x^3 + y^3 = xy. \end{array}$$

9. Show that at points where the radius of curvature is a maximum or minimum the contact of the osculating circle with the curve must be of at least the third order and must always be of odd order.

10. Let PN be a normal to a curve and $P'N$ a neighboring normal. If O is the center of the osculating circle at P, show with the aid of Ex. 6 that ordinarily the perpendicular from O to $P'N$ is of the second order relative to the arc PP' and that the distance ON is of the first order. Hence interpret the statement: Consecutive normals to a curve meet at the center of the osculating circle.

11. Does the osculating circle cross the curve at the point of osculation? Will the osculating circles at neighboring points of the curve intersect in real points?

12. In the hyperbola the focal radii drawn to any point make equal angles with the tangent. Prove this and state and prove the corresponding theorem for the parabola.

13. Given an infinitesimal arc AB cut at C by the perpendicular bisector of its chord AB. What is the order of the difference $AC - BC$?

14. Of what order is the area of the segment included between an infinitesimal arc and its chord compared with the square on the chord?

15. Two sides AB, AC of a triangle are finite and differ infinitesimally; the angle θ at A is an infinitesimal of the same order and the side BC is either rectilinear or curvilinear. What is the order of the neglected infinitesimal if the area is assumed as $\frac{1}{2}\overline{AB}^2\theta$? What if the assumption is $\frac{1}{2}AB\cdot AC\cdot\theta$?

16. A cycloid is the locus of a fixed point upon a circumference which rolls on a straight line. Show that the tangent and normal to the cycloid pass through the highest and lowest points of the rolling circle at each of its instantaneous positions.

17. Show that the increment of arc Δs in the cycloid differs from $2a\sin\frac{1}{2}\theta d\theta$ by an infinitesimal of higher order and that the increment of area (between two consecutive normals) differs from $3a^2\sin^2\frac{1}{2}\theta d\theta$ by an infinitesimal of higher order. Hence show that the total length and area are $8a$ and $3\pi a^2$. Here a is the radius of the generating circle and θ is the angle subtended at the center by the lowest point and the fixed point which traces the cycloid.

18. Show that the radius of curvature of the cycloid is bisected at the lowest point of the generating circle and hence is $4a\sin\frac{1}{2}\theta$.

19. A triangle ABC is circumscribed about any oval curve. Show that if the side BC is bisected at the point of contact, the area of the triangle will be changed by an infinitesimal of the second order when BC is replaced by a neighboring tangent $B'C'$, but that if BC be not bisected, the change will be of the first order. Hence infer that the minimum triangle circumscribed about an oval will have its three sides bisected at the points of contact.

20. If a string is wrapped about a circle of radius a and then unwound so that its end describes a curve, show that the length of the curve and the area between the curve, the circle, and the string are

$$s = \int_0^\theta a\theta d\theta, \qquad A = \int_0^\theta \tfrac{1}{2}a^2\theta^2 d\theta,$$

where θ is the angle that the unwinding string has turned through.

21. Show that the motion in space of a rigid body one point of which is fixed may be regarded as an instantaneous rotation about some axis through the given point. To do this examine the displacements of a unit sphere surrounding the fixed point as center.

22. Suppose a fluid of variable density $D(x)$ is flowing at a given instant through a tube surrounding the x-axis. Let the velocity of the fluid be a function $v(x)$ of x. Show that during the infinitesimal time δt the diminution of the amount of the fluid which lies between $x = a$ and $x = a + h$ is

$$S[v(a+h)D(a+h)\delta t - v(a)D(a)\delta t],$$

where S is the cross section of the tube. Hence show that $D(x)v(x) = \text{const.}$ is the condition that the flow of the fluid shall not change the density at any point.

23. Consider the curve $y = f(x)$ and three equally spaced ordinates at $x = a - \delta$, $x = a$, $x = a + \delta$. Inscribe a trapezoid by joining the ends of the ordinates at $x = a \pm \delta$ and circumscribe a trapezoid by drawing the tangent at the end of the ordinate at $x = a$ and producing to meet the other ordinates. Show that

$$S_0 = 2\delta f(a), \qquad S = 2\delta\left[f(a) + \frac{\delta^2}{6}f''(a) + \frac{\delta^4}{120}f^{\text{iv}}(\xi)\right],$$

$$S_1 = 2\delta\left[f(a) + \frac{\delta^2}{2}f''(a) + \frac{\delta^4}{24}f^{\text{iv}}(\xi_1)\right]$$

are the areas of the circumscribed trapezoid, the curve, the inscribed trapezoid. Hence infer that to compute the area under the curve from the inscribed or circumscribed trapezoids introduces a relative error of the order δ^2, but that to compute from the relation $S = \frac{1}{3}(2S_0 + S_1)$ introduces an error of only the order of δ^4.

24. Let the interval from a to b be divided into an even number $2n$ of equal parts δ and let the $2n+1$ ordinates $y_0, y_1, \cdots, y_{2n}$ at the extremities of the intervals be drawn to the curve $y = f(x)$. Inscribe trapezoids by joining the ends of every other ordinate beginning with y_0, y_2, and going to y_{2n}. Circumscribe trapezoids by drawing tangents at the ends of every other ordinate $y_1, y_3, \cdots, y_{2n-1}$. Compute the area under the curve as

$$S = \int_a^b f(x)\,dx = \frac{b-a}{6n}\left[4(y_1 + y_3 + \cdots + y_{2n-1})\right.$$
$$\left. + 2(y_0 + y_2 + \cdots + y_{2n}) - y_0 - y_{2n}\right] + R$$

by using the work of Ex. 23 and infer that the error R is less than $(b-a)\,\delta^4 f^{(\text{iv})}(\xi)/45$. This method of computation is known as *Simpson's Rule*. It usually gives accuracy sufficient for work to four or even five figures when $\delta = 0.1$ and $b - a = 1$; for $f^{(\text{iv})}(x)$ usually is small.

25. Compute these integrals by Simpson's Rule. Take $2n = 10$ equal intervals. Carry numerical work to six figures except where tables must be used to find $f(x)$:

$(\alpha)\ \displaystyle\int_1^2 \frac{dx}{x} = \log 2 = 0.69315,$ $(\beta)\ \displaystyle\int_0^1 \frac{dx}{1+x^2} = \tan^{-1} 1 = \frac{1}{4}\pi = 0.78535,$

$(\gamma)\ \displaystyle\int_0^{\frac{1}{2}\pi} \sin x\,dx = 1.00000,$ $(\delta)\ \displaystyle\int_1^2 \log_{10} x\,dx = 2\log_{10} x - M = 0.16776,$

$(\epsilon)\ \displaystyle\int_0^1 \frac{\log(1+x)}{1+x^2}\,dx = 0.27220,$ $(\zeta)\ \displaystyle\int_0^1 \frac{\log(1+x)}{x}\,dx = 0.82247.$

The answers here given are the true values of the integrals to five places.

26. Show that the quadrant of the ellipse $x = a\sin\phi$, $y = b\cos\phi$ is

$$s = a\int_0^{\frac{1}{2}\pi} \sqrt{1 - e^2\sin^2\phi}\,d\phi = \tfrac{1}{2}\pi a\int_0^1 \sqrt{\tfrac{1}{2}(2 - e^2) + \tfrac{1}{2}e^2\cos\pi u}\,du.$$

Compute to four figures by Simpson's Rule with six divisions the quadrants of the ellipses:

$$(\alpha)\ e = \tfrac{1}{2}\sqrt{3},\quad s = 1.211\,a, \qquad (\beta)\ e = \tfrac{1}{2}\sqrt{2},\quad s = 1.351\,a.$$

27. Expand s in Ex. 26 into a series and discuss the remainder.

$$s = \frac{1}{2}\pi a\left[1 - \left(\frac{1}{2}\right)^2 e^2 - \left(\frac{1\cdot 3}{2\cdot 4}\right)^2\frac{e^4}{3} - \left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2\frac{e^6}{5} - \cdots - \left(\frac{1\cdot 3\cdots(2n-1)}{2\cdot 4\cdots 2n}\right)^2\frac{e^{2n}}{2n-1} - R_n\right]$$

$$R_n < \frac{1}{1-e^2}\left(\frac{1\cdot 3\cdots(2n+1)}{2\cdot 4\cdots(2n+2)}\right)^2\frac{e^{2n+2}}{2n+1} \qquad \text{See Ex. 18, p. 60, and Peirce's "Tables," p. 62.}$$

Estimate the number of terms necessary to compute Ex. 26 (β) with an error not greater than 2 in the last place and compare the labor with that of Simpson's Rule.

28. If the eccentricity of an ellipse is $\frac{1}{10}$, find to five decimals the percentage error made in taking $2\pi a$ as the perimeter. *Ans.* 0.00691 %.

29. If the catenary $y = c \cosh(x/c)$ gives the shape of a wire of length L suspended between two points at the same level and at a distance l nearly equal to L, find the first approximation connecting L, l, and d, where d is the dip of the wire at its lowest point below the level of support.

30. At its middle point the parabolic cable of a suspension bridge 1000 ft. long between the supports sags 50 ft. below the level of the ends. Find the length of the cable correct to inches.

40. Some differential geometry. Suppose that between the increments of a set of variables all of which depend on a single variable t there exists an equation which is true except for infinitesimals of higher order than $\Delta t = dt$, then the equation will be exactly true for the differentials of the variables. Thus if

$$f\Delta x + g\Delta y + h\Delta z + l\Delta t + \cdots + e_1 + e_2 + \cdots = 0$$

is an equation of the sort mentioned and if the coefficients are any functions of the variables and if $e_1, e_2, \cdots$ are infinitesimals of higher order than dt, the limit of

$$f\frac{\Delta x}{\Delta t} + g\frac{\Delta y}{\Delta t} + h\frac{\Delta z}{\Delta t} + l\frac{\Delta t}{\Delta t} + \cdots + \frac{e_1}{\Delta t} + \frac{e_2}{\Delta t} = 0$$

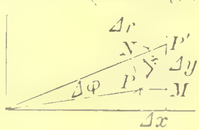
is
$$f\frac{dx}{dt} + g\frac{dy}{dt} + h\frac{dz}{dt} + l = 0,$$

or
$$f dx + g dy + h dz + l dt = 0;$$

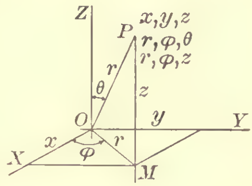
and the statement is proved. This result is very useful in writing down various differential formulas of geometry where the approximate relation between the increments is obvious and where the true relation between the differentials can therefore be found.

For instance in the case of the differential of arc in rectangular coördinates, if the increment of arc is known to differ from its chord by an infinitesimal of higher order, the Pythagorean theorem shows that the equation
$$\Delta s^2 = \Delta x^2 + \Delta y^2 \quad \text{or} \quad \Delta s^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 \tag{7}$$
is true except for infinitesimals of higher order; and hence
$$ds^2 = dx^2 + dy^2 \quad \text{or} \quad ds^2 = dx^2 + dy^2 + dz^2. \tag{7'}$$

In the case of plane polar coördinates, the triangle $PP'N$ (see Fig.) has two curvilinear sides PP' and PN and is right-angled at N. The Pythagorean theorem may be applied to a curvilinear triangle, or the triangle may be replaced by the rectilinear triangle $PP'N$ with the angle at N no longer a right angle but nearly so. In either way of looking at the figure, it is easily seen that the equation $\Delta s^2 = \Delta r^2 + r^2\Delta\phi^2$



which the figure suggests differs from a true equation by an infinitesimal of higher order; and hence the inference that in polar coördinates $ds^2 = dr^2 + r^2 d\phi^2$.

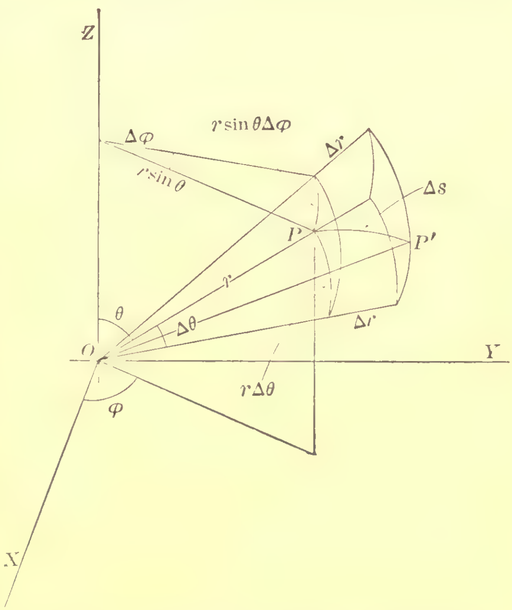


The two most used systems of coördinates other than rectangular in space are the *polar* or *spherical* and the *cylindrical*. In the first the distance $r = OP$ from the pole or center, the longitude or meridional angle ϕ, and the colatitude or polar angle θ are chosen as coördinates; in the second, ordinary polar coördinates $r = OM$ and ϕ in the xy-plane are combined with the ordinary rectangular z for distance from that plane. The formulas of transformation are

$$\begin{aligned} z &= r\cos\theta, & r &= \sqrt{x^2+y^2+z^2}, \\ y &= r\sin\theta\sin\phi, & \theta &= \cos^{-1}\frac{z}{\sqrt{x^2+y^2+z^2}}, \\ x &= r\sin\theta\cos\phi, & \phi &= \tan^{-1}\frac{y}{x}, \end{aligned} \qquad (8)$$

for polar coördinates, and for cylindrical coördinates they are

$$z = z, \quad y = r\sin\phi, \quad x = r\cos\phi, \quad r = \sqrt{x^2+y^2}, \quad \phi = \tan^{-1}\frac{y}{x}. \qquad (9)$$



Formulas such as that for the differential of arc may be obtained for these new coördinates by mere transformation of (7′) according to the rules for change of variable.

In both these cases, however, the value of ds may be found readily by direct inspection of the figure. The small parallelepiped (figure for polar case) of which Δs is the diagonal has some of its edges and faces curved instead of straight; all the angles, however, are right angles, and as the edges are infinitesimal, the equations certainly suggested as holding except for infinitesimals of higher order are

$$\Delta s^2 = \Delta r^2 + r^2 \sin^2\theta \Delta\phi^2 + r^2\Delta\theta^2 \quad \text{and} \quad \Delta s^2 = \Delta r^2 + r^2\Delta\phi^2 + \Delta z^2 \quad (10)$$

or $$ds^2 = dr^2 + r^2 \sin^2\theta d\phi^2 + r^2 d\theta^2 \quad \text{and} \quad ds^2 = dr^2 + r^2 d\phi^2 + dz^2. \quad (10')$$

To make the proof complete, it would be necessary to show that nothing but infinitesimals of higher order have been neglected and it might actually be easier to transform $\sqrt{dx^2 + dy^2 + dz^2}$ rather than give a rigorous demonstration of this fact. Indeed the infinitesimal method is seldom used rigorously; its great use is to make the facts so clear to the rapid worker that he is willing to take the evidence and omit the proof.

In the plane for rectangular coördinates with rulings parallel to the y-axis and for polar coördinates with rulings issuing from the pole the increments of area differ from

$$dA = y dx \quad \text{and} \quad dA = \tfrac{1}{2} r^2 d\phi \qquad (11)$$

respectively by infinitesimals of higher order, and

$$A = \int_{x_0}^{x_1} y dx \quad \text{and} \quad A = \tfrac{1}{2}\int_{\phi_0}^{\phi_1} r^2 d\phi \qquad (11')$$

are therefore the formulas for the area under a curve and between two ordinates, and for the area between the curve and two radii. If the plane is ruled by lines parallel to both axes or by lines issuing from the pole and by circles concentric with the pole, as is customary for double integration (§§ 131, 134), the increments of area differ respectively by infinitesimals of higher order from

$$dA = dx dy \quad \text{and} \quad dA = r dr d\phi, \qquad (12)$$

and the formulas for the area in the two cases are

$$A = \lim \sum \Delta A = \iint dA = \iint dx dy, \qquad (12')$$

$$A = \lim \sum \Delta A = \iint dA = \iint r dr d\phi,$$

where the double integrals are extended over the area desired.

The elements of volume which are required for triple integration (§§ 133, 134) over a volume in space may readily be written down for the three cases of rectangular, polar, and cylindrical coördinates. In the first case space is supposed to be divided up by planes $x = a$, $y = b$, $z = c$ perpendicular to the axes and spaced at infinitesimal intervals; in the second case the division is made by the spheres $r = a$ concentric with the pole, the planes $\phi = b$ through the polar axis, and the cones $\theta = c$ of revolution about the polar axis; in the third case by the cylinders $r = a$, the planes $\phi = b$, and the planes $z = c$. The infinitesimal

volumes into which space is divided then differ from

$$dv = dxdydz, \qquad dv = r^2 \sin\theta dr d\phi d\theta, \qquad dv = r dr d\phi dz \tag{13}$$

respectively by infinitesimals of higher order, and

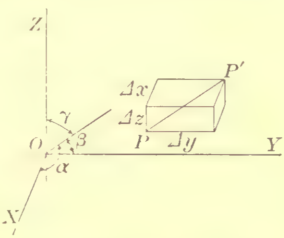
$$\iiint dxdydz, \qquad \iiint r^2 \sin\theta dr d\phi d\theta, \qquad \iiint r dr d\phi dz \tag{13'}$$

are the formulas for the volumes.

41. The direction of a line in space is represented by the three angles which the line makes with the positive directions of the axes or by the cosines of those angles, the direction cosines of the line. From the definition and figure it appears that

$$l = \cos\alpha = \frac{dx}{ds}, \qquad m = \cos\beta = \frac{dy}{ds}, \qquad n = \cos\gamma = \frac{dz}{ds} \tag{14}$$

are the direction cosines of the tangent to the arc at the point; of the tangent and not of the chord for the reason that the increments are replaced by the differentials. Hence it is seen that for the *direction cosines of the tangent* the proportion



$$l : m : n = dx : dy : dz \tag{14'}$$

holds. The equations of a space curve are

$$x = f(t), \qquad y = g(t), \qquad z = h(t)$$

in terms of a variable parameter t.* At the point (x_0, y_0, z_0) where $t = t_0$ the *equations of the tangent lines* would then be

$$\frac{x - x_0}{(dx)_0} = \frac{y - y_0}{(dy)_0} = \frac{z - z_0}{(dz)_0} \quad \text{or} \quad \frac{x - x_0}{f'(t_0)} = \frac{y - y_0}{g'(t_0)} = \frac{z - z_0}{h'(t_0)}. \tag{15}$$

As the cosine of the angle θ between the two directions given by the direction cosines l, m, n and l', m', n' is

$$\cos\theta = ll' + mm' + nn', \quad \text{so} \quad ll' + mm' + nn' = 0 \tag{16}$$

is the condition for the perpendicularity of the lines. Now if (x, y, z) lies in the plane normal to the curve at x_0, y_0, z_0, the lines determined by the ratios $x - x_0 : y - y_0 : z - z_0$ and $(dx)_0 : (dy)_0 : (dz)_0$ will be perpendicular. Hence the *equation of the normal plane* is

$$(x - x_0)(dx)_0 + (y - y_0)(dy)_0 + (z - z_0)(dz)_0 = 0$$

or

$$f'(t_0)(x - x_0) + g'(t_0)(y - y_0) + h'(t_0)(z - z_0) = 0. \tag{17}$$

* For the sake of generality the parametric form in t is assumed; in a particular case a simplification might be made by taking one of the variables as t and one of the functions f', g', h' would then be 1. Thus in Ex. 8 (e), y should be taken as t.

The *tangent plane* to the curve is not determinate; any plane through the tangent line will be tangent to the curve. If λ be a parameter, the pencil of tangent planes is

$$\frac{x - x_0}{f'(t_0)} + \lambda \frac{y - y_0}{g'(t_0)} - (1 + \lambda) \frac{z - z_0}{h'(t_0)} = 0.$$

There is one particular tangent plane, called *the osculating plane*, which is of especial importance. Let

$$x - x_0 = f'(t_0)\tau + \tfrac{1}{2} f''(t_0)\tau^2 + \tfrac{1}{6} f'''(\xi)\tau^3, \quad \tau = t - t_0, \quad t_0 < \xi < t,$$

with similar expansions for y and z, be the Taylor developments of x, y, z about the point of tangency. When these are substituted in the equation of the plane, the result is

$$\frac{1}{2}\tau^2\left[\frac{f''(t_0)}{f'(t_0)} + \lambda \frac{g''(t_0)}{g'(t_0)} - (1 + \lambda) \frac{h''(t_0)}{h'(t_0)}\right] + \frac{1}{6}\tau^3\left[\frac{f'''(\xi)}{f'(t_0)} + \lambda \frac{g'''(\eta)}{g'(t_0)} - (1 + \lambda) \frac{h'''(\zeta)}{h'(t_0)}\right].$$

This expression is of course proportional to the distance from any point x, y, z of the curve to the tangent plane and is seen to be in general of the second order with respect to τ or ds. It is, however, possible to choose for λ that value which makes the first bracket vanish. The tangent plane thus selected has the property that *the distance of the curve from it in the neighborhood of the point of tangency is of the third order and is called the osculating plane.* The substitution of the value of λ gives

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ f'(t_0) & g'(t_0) & h'(t_0) \\ f''(t_0) & g''(t_0) & h''(t_0) \end{vmatrix} = 0 \quad \text{or} \quad \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ (dx)_0 & (dy)_0 & (dz)_0 \\ (d^2x)_0 & (d^2y)_0 & (d^2z)_0 \end{vmatrix} = 0 \qquad (18)$$

or
$$(dyd^2z - dzd^2y)_0(x - x_0) + (dzd^2x - dxd^2z)_0(y - y_0) + (dxd^2y - dyd^2x)_0(z - z_0) = 0$$

as the equation of the osculating plane. In case $f''(t_0) = g''(t_0) = h''(t_0) = 0$, this equation of the osculating plane vanishes identically and it is necessary to push the development further (Ex. 11).

42. For the case of plane curves the *curvature* is defined as the rate at which the tangent turns compared with the description of arc, that is, as $d\phi/ds$ if $d\phi$ denotes the differential of the angle through which the tangent turns when the point of tangency advances along the curve by ds. The radius of curvature R is the reciprocal of the curvature, that is, it is $ds/d\phi$. Then

$$d\phi = d\tan^{-1}\frac{dy}{dx}, \qquad \frac{d\phi}{ds} = \frac{d\phi}{dx}\frac{dx}{ds} = \frac{y''}{[1 + y'^2]^{\frac{3}{2}}}, \qquad R = \frac{[1 + y'^2]^{\frac{3}{2}}}{y''}. \quad (19)$$

where accents denote differentiation with respect to x. For space curves the same definitions are given. If l, m, n and $l+dl$, $m+dm$, $n+dn$ are the direction cosines of two successive tangents,

$$\cos d\phi = l(l+dl) + m(m+dm) + n(n+dn).$$

But $\quad l^2+m^2+n^2 = 1 \quad$ and $\quad (l+dl)^2 + (m+dm)^2 + (n+dn)^2 = 1.$

Hence $\quad dl^2 + dm^2 + dn^2 = 2 - 2\cos d\phi = (2\sin\frac{1}{2}d\phi)^2,$

$$\frac{1}{R^2} = \left(\frac{d\phi}{ds}\right)^2 = \left[\frac{2\sin\frac{1}{2}d\phi}{ds}\right]^2 = \frac{dl^2+dm^2+dn^2}{ds^2} = l'^2 + m'^2 + n'^2, \quad (19')$$

where accents denote differentiation with respect to s.

The *torsion* of a space curve is defined as the rate of turning of the osculating plane compared with the increase of arc (that is, $d\psi/ds$, where $d\psi$ is the differential angle the normal to the osculating plane turns through), and may clearly be calculated by the same formula as the curvature provided the direction cosines L, M, N of the normal to the plane take the places of the direction cosines l, m, n of the tangent line. Hence the torsion is

$$\frac{1}{\mathrm{R}^2} = \left(\frac{d\psi}{ds}\right)^2 = \frac{dL^2+dM^2+dN^2}{ds^2} = L'^2 + M'^2 + N'^2; \qquad (20)$$

and the radius of torsion R is defined as the reciprocal of the torsion, where from the equation of the osculating plane

$$\frac{L}{dyd^2z - dzd^2y} = \frac{M}{dzd^2x - dxd^2z} = \frac{N}{dxd^2y - dyd^2x}$$
$$= \frac{1}{\sqrt{\text{sum of squares}}}. \qquad (20')$$

The actual computation of these quantities is somewhat tedious.

The vectorial discussion of curvature and torsion (§ 77) gives a better insight into the principal directions connected with a space curve. These are the direction of the *tangent*, that of the normal in the osculating plane and directed towards the concave side of the curve and called the *principal normal*, and that of the normal to the osculating plane drawn upon that side which makes the three directions form a right-handed system and called the *binormal*. In the notations there given, combined with those above,

$$\mathbf{r} = x\mathbf{i} + y\mathbf{i} + z\mathbf{k}, \qquad \mathbf{t} = l\mathbf{i} + m\mathbf{j} + n\mathbf{k}, \qquad \mathbf{c} = \lambda\mathbf{i} + \mu\mathbf{j} + \nu\mathbf{k}, \qquad \mathbf{n} = L\mathbf{i} + M\mathbf{j} + N\mathbf{k},$$

where λ, μ, ν are taken as the direction cosines of the principal normal. Now $d\mathbf{t}$ is parallel to $\mathbf{c}$ and $d\mathbf{n}$ is parallel to $-\mathbf{c}$. Hence the results

$$\frac{dl}{\lambda} = \frac{dm}{\mu} = \frac{dn}{\nu} = \frac{ds}{R} \quad \text{and} \quad \frac{dL}{\lambda} = \frac{dM}{\mu} = \frac{dN}{\nu} = -\frac{ds}{\mathrm{R}} \qquad (21)$$

follow from $d\mathbf{t}/ds = \mathbf{C}$ and $d\mathbf{n}/ds = \mathbf{T}$. Now $d\mathbf{c}$ is perpendicular to $\mathbf{c}$ and hence in the plane of $\mathbf{t}$ and $\mathbf{n}$; it may be written as $d\mathbf{c} = (\mathbf{t} \cdot d\mathbf{c})\mathbf{t} + (\mathbf{n} \cdot d\mathbf{c})\mathbf{n}$. But as $\mathbf{t} \cdot \mathbf{c} = \mathbf{n} \cdot \mathbf{c} = 0$, $\mathbf{t} \cdot d\mathbf{c} = -\mathbf{c} \cdot d\mathbf{t}$ and $\mathbf{n} \cdot d\mathbf{c} = -\mathbf{c} \cdot d\mathbf{n}$. Hence

$$d\mathbf{c} = -(\mathbf{c} \cdot d\mathbf{t})\mathbf{t} - (\mathbf{c} \cdot d\mathbf{n})\mathbf{n} = -C\mathbf{t}\,ds + T\mathbf{n}\,ds = -\frac{\mathbf{t}}{R}ds + \frac{\mathbf{n}}{\mathrm{R}}ds.$$

Hence $$\frac{d\lambda}{ds} = -\frac{l}{R} + \frac{L}{\mathrm{R}}, \qquad \frac{d\mu}{ds} = -\frac{m}{R} + \frac{M}{\mathrm{R}}, \qquad \frac{d\nu}{ds} = -\frac{n}{R} + \frac{N}{\mathrm{R}}. \tag{22}$$

Formulas (22) are known as *Frenet's Formulas*; they are usually written with $-\mathrm{R}$ in the place of R because a left-handed system of axes is used and the torsion, being an odd function, changes its sign when all the axes are reversed. If accents denote differentiation by s,

$$\text{above formulas, right-handed } \frac{1}{\mathrm{R}} = \frac{\begin{vmatrix} x' & y' & z' \\ x'' & y'' & z'' \\ x''' & y''' & z''' \end{vmatrix}}{x''^2 + y''^2 + z''^2}. \quad \text{usual formulas, left-handed } \frac{1}{\mathrm{R}} = -\frac{\begin{vmatrix} x' & y' & z' \\ x'' & y'' & z'' \\ x''' & y''' & z''' \end{vmatrix}}{x''^2 + y''^2 + z''^2}. \tag{23}$$

EXERCISES

1. Show that in polar coördinates in the plane, the tangent of the inclination of the curve to the radius vector is $rd\phi/dr$.

2. Verify (10), (10′) by direct transformation of coördinates.

3. Fill in the steps omitted in the text in regard to the proof of (10), (10′) by the method of infinitesimal analysis.

4. A rhumb line on a sphere is a line which cuts all the meridians at a constant angle, say α. Show that for a rhumb line $\sin\theta d\phi = \tan\alpha d\theta$ and $ds = r\sec\alpha d\theta$. Hence find the equation of the line, show that it coils indefinitely around the poles of the sphere, and that its total length is $\pi r\sec\alpha$.

5. Show that the surfaces represented by $F(\phi, \theta) = 0$ and $F(r, \theta) = 0$ in polar coördinates in space are respectively cones and surfaces of revolution about the polar axis. What sort of surface would the equation $F(r, \phi) = 0$ represent?

6. Show accurately that the expression given for the differential of area in polar coördinates in the plane and for the differentials of volume in polar and cylindrical coördinates in space differ from the corresponding increments by infinitesimals of higher order.

7. Show that $\dfrac{dr}{ds}$, $r\dfrac{d\theta}{ds}$, $r\sin\theta\dfrac{d\phi}{ds}$ are the direction cosines of the tangent to a space curve relative to the radius, meridian, and parallel of latitude.

8. Find the tangent line and normal plane of these curves.

(α) $xyz = 1$, $y^2 = x$ at $(1, 1, 1)$. (β) $x = a\cos t$, $y = a\sin t$, $z = kt$.

(γ) $2ay = x^2$, $6a^2z = x^3$. (δ) $x = t\cos t$, $y = t\sin t$, $z = kt$.

(ϵ) $y = x^2$, $z^2 = 1 - y$. (ζ) $x^2 + y^2 + z^2 = a^2$, $x^2 - y^2 - 2ax = 0$.

9. Find the equation of the osculating plane in the examples of Ex. 8. Note that if x is the independent variable, the equation of the plane is

$$\left(\frac{dy}{dx}\frac{d^2z}{dx^2} - \frac{dz}{dx}\frac{d^2y}{dx^2}\right)_0 (x - x_0) - \left(\frac{d^2z}{dx^2}\right)_0 (y - y_0) + \left(\frac{d^2y}{dx^2}\right)_0 (z - z_0) = 0.$$

10. A space curve passes through the origin, is tangent to the x-axis, and has $z = 0$ as its osculating plane at the origin. Show that

$$x = tf'(0) + \tfrac{1}{2}t^2f''(0) + \cdots, \qquad y = \tfrac{1}{2}t^2g''(0) + \cdots, \qquad z = \tfrac{1}{6}t^3h'''(0) + \cdots$$

will be the form of its Maclaurin development if $t = 0$ gives $x = y = z = 0$.

11. If the 2d, 3d, $\cdots$, $(n-1)$st derivatives of f, g, h vanish for $t = t_0$ but not all the nth derivatives vanish, show that there is a plane from which the curve departs by an infinitesimal of the $(n+1)$st order and with which it therefore has contact of order n. Such a plane is called a hyperosculating plane. Find its equation.

12. At what points if any do the curves (β), (γ), (ϵ), (ζ), Ex. 8 have hyperosculating planes and what is the degree of contact in each case?

13. Show that the expression for the radius of curvature is

$$\frac{1}{R} = \sqrt{x''^2 + y''^2 + z''^2} = \frac{[(g'h'' - h'g'')^2 + (h'f'' - f'h'')^2 + (f'g'' - g'f'')^2]^{\frac{1}{2}}}{[f'^2 + g'^2 + h'^2]^{\frac{3}{2}}},$$

where in the first case accents denote differentiation by s, in the second by t.

14. Show that the radius of curvature of a space curve is the radius of curvature of its projection on the osculating plane at the point in question.

15. From Frenet's Formulas show that the successive derivatives of x are

$$x' = l, \qquad x'' = l' = \frac{\lambda}{R}, \qquad x''' = \frac{\lambda'}{R} - \frac{\lambda R'}{R^2} = -\frac{l}{R^2} - \lambda\frac{R'}{R^2} + \frac{L}{R\mathrm{R}},$$

where accents denote differentiation by s. Show that the results for y and z are the same except that m, μ, M or n, ν, N take the places of l, λ, L. Hence infer that for the nth derivatives the results are

$$x^{(n)} = lP_1 + \lambda P_2 + LP_3, \qquad y^{(n)} = mP_1 + \mu P_2 + MP_3, \qquad z^{(n)} = nP_1 + \nu P_2 + NP_3,$$

where P_1, P_2, P_3 are rational functions of R and R and their derivatives by s.

16. Apply the foregoing to the expansion of Ex. 10 to show that

$$x = s - \frac{1}{6R^2}s^3 + \cdots, \qquad y = \frac{s^2}{2R} - \frac{R'}{6R^2}s^3 + \cdots, \qquad z = \frac{s^3}{6R\mathrm{R}} + \cdots,$$

where R and R are the values at the origin where $s = 0$, $l = \mu = N = 1$, and the other six direction cosines m, n, λ, ν, L, M vanish. Find s and write the expansion of the curve of Ex. 8 (γ) in this form.

17. Note that the distance of a point on the curve as expanded in Ex. 16 from the sphere through the origin and with center at the point $(0, R, R'\mathrm{R})$ is

$$\sqrt{x^2 + (y - R)^2 + (z - R'\mathrm{R})^2} - \sqrt{R^2 + R'^2\mathrm{R}^2}$$
$$= \frac{(x^2 + y^2 - 2Ry + z^2 - 2R'\mathrm{R}z)}{\sqrt{x^2 + (y - R)^2 + (z - R'\mathrm{R})^2} + \sqrt{R^2 + R'^2\mathrm{R}^2}},$$

and consequently is of the fourth order. The curve therefore has contact of the third order with this sphere. Can the equation of this sphere be derived by a limiting process like that of Ex. 18 as applied to the osculating plane

18. The osculating plane may be regarded as the plane passed through three consecutive points of the curve; in fact it is easily shown that

$$\lim_{\substack{\delta x,\, \delta y,\, \delta z \\ \Delta x,\, \Delta y,\, \Delta z \\ \text{approach } 0}} \begin{vmatrix} x & y & z & 1 \\ x_0 & y_0 & z_0 & 1 \\ x_0 + \delta x & y_0 + \delta y & z_0 + \delta z & 1 \\ x_0 + \Delta x & y_0 + \Delta y & z_0 + \Delta z & 1 \end{vmatrix} = \frac{1}{2} \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ (dx)_0 & (dy)_0 & (dz)_0 \\ (d^2x)_0 & (d^2y)_0 & (d^2z)_0 \end{vmatrix} = 0.$$

19. Express the radius of torsion in terms of the derivatives of x, y, z by t (Ex. 10, p. 67).

20. Find the direction, curvature, osculating plane, torsion, and osculating sphere (Ex. 17) of the conical helix $x = t\cos t$, $y = t\sin t$, $z = kt$ at $t = 2\pi$.

21. Upon a plane diagram which shows Δs, Δx, Δy, exhibit the lines which represent ds, dx, dy under the different hypotheses that x, y, or s is the independent variable.

CHAPTER IV

PARTIAL DIFFERENTIATION; EXPLICIT FUNCTIONS

43. Functions of two or more variables. The definitions and theorems about functions of more than one independent variable are to a large extent similar to those given in Chap. II for functions of a single variable, and the changes and difficulties which occur are for the most part amply illustrated by the case of two variables. The work in the text will therefore be confined largely to this case and the generalizations to functions involving more than two variables may be left as exercises.

If the value of a variable z is uniquely determined when the values (x, y) of two variables are known, z is said to be a function $z = f(x, y)$ of the two variables. The set of values $[(x, y)]$ or of points $P(x, y)$ of the xy-plane for which z is defined may be any set, but usually consists of all the points in a certain area or region of the plane bounded by a curve which may or may not belong to the region, just as the end points of an interval may or may not belong to it. Thus the function $1/\sqrt{1 - x^2 - y^2}$ is defined for all points within the circle $x^2 + y^2 = 1$, but not for points on the perimeter of the circle. For most purposes it is sufficient to think of the boundary of the region of definition as a polygon whose sides are straight lines or such curves as the geometric intuition naturally suggests.

The first way of representing the function $z = f(x, y)$ geometrically is by *the surface* $z = f(x, y)$, just as $y = f(x)$ was represented by a curve. This method is not available for $u = f(x, y, z)$, a function of three variables, or for functions of a greater number of variables; for space has only three dimensions. A second method of representing the function $z = f(x, y)$ is by its *contour lines* in the xy-plane, that is, the curves $f(x, y) =$ const. are plotted and to each curve is attached the value of the constant. This is the method employed on maps in marking heights above sea level or depths of the ocean below sea level. It is evident that these contour lines are nothing but the projections on the xy-plane of the curves in which the surface $z = f(x, y)$ is cut by the planes $z =$ const. This method is applicable to functions $u = f(x, y, z)$ of three variables. The *contour surfaces* $u =$ const. which are thus obtained

are frequently called *equipotential surfaces*. If the function is single valued, the contour lines or surfaces cannot intersect one another.

The *function* $z = f(x, y)$ *is continuous for* (a, b) when either of the following equivalent conditions is satisfied:

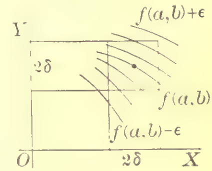
1°. $\lim f(x, y) = f(a, b)$ *or* $\lim f(x, y) = f(\lim x, \lim y)$,

no matter how the variable point $P(x, y)$ *approaches* (a, b).

2°. *If for any assigned* ϵ, *a number* δ *may be found so that*

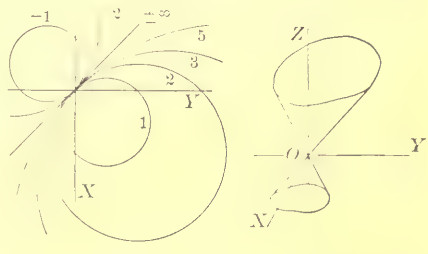
$$|f(x, y) - f(a, b)| < \epsilon \quad \textit{when} \quad |x - a| < \delta, \ |y - b| < \delta.$$

Geometrically this means that if a square with (a, b) as center and with sides of length 2δ parallel to the axes be drawn, the portion of the surface $z = f(x, y)$ above the square will lie between the two planes $z = f(a, b) \pm \epsilon$. Or if contour lines are used, no line $f(x, y) =$ const. where the constant differs from $f(a, b)$ by so much as ϵ will cut into the square. It is clear that in place of a square surrounding (a, b) a circle of radius δ or any other figure which lay within the square might be used.



44. *Continuity examined.* From the definition of continuity just given and from the corresponding definition in § 24, it follows that if $f(x, y)$ is a continuous function of x and y for (a, b), then $f(x, b)$ is a continuous function of x for $x = a$ and $f(a, y)$ is a continuous function of y for $y = b$. That is, if f is continuous in x and y jointly, it is continuous in x and y severally. It might be thought that conversely if $f(x, b)$ is continuous for $x = a$ and $f(a, y)$ for $y = b$, $f(x, y)$ would be continuous in (x, y) for (a, b). That is, if f is continuous in x and y severally, it would be continuous in x and y jointly. A simple example will show that this is *not necessarily true*. Consider the case

$$z = f(x, y) = \frac{x^2 + y^2}{x + y}, \qquad f(0, 0) = 0$$



and examine z for continuity at $(0, 0)$. The functions $f(x, 0) = x$, and $f(0, y) = y$ are surely continuous in their respective variables. But the surface $z = f(x, y)$ is a conical surface (except for the points of the z-axis other than the origin) and it is clear that $P(x, y)$ may approach the origin in such a manner that z shall approach any desired value. Moreover, a glance at the contour lines shows that they all enter any circle or square, no matter how small, concentric with the origin. If P approaches the origin along one of these lines, z remains constant and its limiting value is that constant. In fact by approaching the origin along a set of points which jump from one contour line to another, a method of approach may be found such that z approaches no limit whatsoever but oscillates between wide limits or becomes infinite. Clearly the conditions of continuity are not at all fulfilled by z at $(0, 0)$.

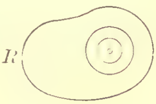
Double limits. There often arise for consideration expressions like

$$\lim_{y \doteq b}\left[\lim_{x \doteq a} f(x, y)\right], \qquad \lim_{x \doteq a}\left[\lim_{y \doteq b} f(x, y)\right], \tag{1}$$

where the limits exist whether x first approaches its limit, and then y its limit, or vice versa, and where the question arises as to whether the two limits thus obtained are equal, that is, whether the order of taking the limits in the double limit may be interchanged. It is clear that if the function $f(x, y)$ is continuous at (a, b), the limits approached by the two expressions will be equal; for the limit of $f(x, y)$ is $f(a, b)$ no matter how (x, y) approaches (a, b). If f is discontinuous at (a, b), it may still happen that the order of the limits in the double limit may be interchanged, as was true in the case above where the value in either order was zero; but this cannot be affirmed in general, and special considerations must be applied to each case when f is discontinuous.

*Varieties of regions.** For both pure mathematics and physics the classification of regions according to their *connectivity* is important. Consider a finite region R bounded by a curve which nowhere cuts itself. (For the present purposes it is not necessary to enter upon the subtleties of the meaning of "curve" (see §§ 127–128); ordinary intuition will suffice.) It is clear that if any closed curve drawn in this region had an unlimited tendency to contract, it could draw together to a point and disappear. On the other hand, if R' be a region like R except that a portion has been removed so that R' is bounded by two curves one within the other, it is clear that some closed curves, namely those which did not encircle the portion removed, could shrink away to a point, whereas other closed curves, namely those which encircled that portion, could at most shrink down into coincidence with the boundary of that portion. Again, if two portions are removed so as to give rise to the region R'', there are circuits around each of the portions which at most can only shrink down to the boundaries of those portions and circuits around both portions which can shrink down to the boundaries and a line joining them. A region like R, where *any* closed curve or circuit may be shrunk away to nothing is called a *simply connected region;* whereas regions in which there are circuits which cannot be shrunk away to nothing are called *multiply connected* regions.

R



R'

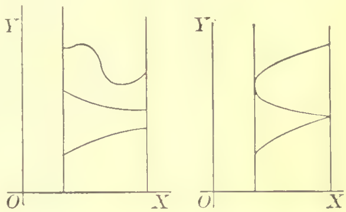
R''

A multiply connected region may be made simply connected by a simple device and convention. For suppose that in R' a line were drawn connecting the two bounding curves and it were agreed that no curve or circuit drawn within R' should cross this line. Then the entire region would be surrounded by a single boundary, part of which would be counted twice. The figure indicates the situation. In like manner if two lines were drawn in R'' connecting both interior boundaries to the exterior or connecting the two interior boundaries together and either of them to the outer boundary, the region would be rendered simply connected. The entire region would have a single boundary of which parts would be counted twice, and any circuit which did not cross the lines could be shrunk away to nothing. The lines

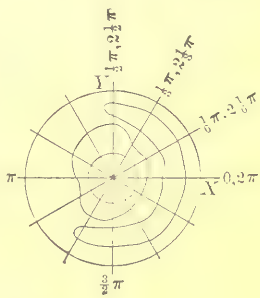
* The discussion from this point to the end of § 45 may be connected with that of §§ 123–126.

thus drawn in the region to make it simply connected are called *cuts*. There is no need that the region be finite; it might extend off indefinitely in some directions like the region between two parallel lines or between the sides of an angle, or like the entire half of the xy-plane for which y is positive. In such cases the cuts may be drawn either to the boundary or off indefinitely in such a way as not to meet the boundary.

45. *Multiple valued functions.* If more than one value of z corresponds to the pair of values (x, y), the function z is multiple valued, and there are some noteworthy differences between multiple valued functions of one variable and of several variables. It was stated (§ 23) that multiple valued functions were divided into branches each of which was single valued. There are two cases to consider when there is one variable, and they are illustrated in the figure. Either there is no value of x in the interval for which the different values of the function are equal and there is consequently a number D which gives the least value of the difference between any two branches, or there is a value of x for which different branches have the same value. Now in the first case, if x changes its value continuously and if $f(x)$ be constrained also to change continuously, there is no possibility of passing from one branch of the function to another; but in the second case such change is possible for, when x passes through the value for which the branches have the same value, the function while constrained to change its value continuously may turn off onto the other branch, although it need not do so.



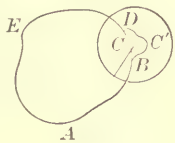
In the case of a function $z = f(x, y)$ of two variables, it is not true that if the values of the function nowhere become equal in or on the boundary of the region over which the function is defined, then it is impossible to pass continuously from one branch to another, and if $P(x, y)$ describes any continuous closed curve or circuit in the region, the value of $f(x, y)$ changing continuously must return to its original value when P has completed the description of the circuit. For suppose the function z be a helicoidal surface $z = a\tan^{-1}(y/x)$, or rather the portion of that surface between two cylindrical surfaces concentric with the axis of the helicoid, as is the case of the surface of the screw of a jack, and the circuit be taken around the inner cylinder. The multiple numbering of the contour lines indicates the fact that the function is multiple valued. Clearly, each time that the circuit is described, the value of z is increased by the amount between the successive branches or leaves of the surface (or decreased by that amount if the circuit is described in the opposite direction). The region here dealt with is not simply connected and the circuit cannot be shrunk to nothing — which is the key to the situation.

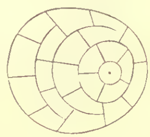


Theorem. If the difference between the different values of a continuous multiple valued function is never less than a finite number D for any set (x, y) of values of the variables whether in or upon the boundary of the region of definition, then the value $f(x, y)$ of the function, constrained to change continuously,

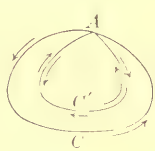
will return to its initial value when the point $P(x, y)$, describing a closed curve which can be shrunk to nothing, completes the circuit and returns to its starting point.

Now owing to the continuity of f throughout the region, it is possible to find a number δ so that $|f(x, y) - f(x', y')| < \epsilon$ when $|x - x'| < \delta$ and $|y - y'| < \delta$ no matter what points of the region (x, y) and (x', y') may be. Hence the values of f at any two points of a small region which lies within any circle of radius $\frac{1}{2}\delta$ cannot differ by so much as the amount D. If, then, the circuit is so small that it may be inclosed within such a circle, there is no possibility of passing from one value of f to another when the circuit is described and f must return to its initial value. Next let there be given any circuit such that the value of f starting from a given value $f(x, y)$ returns to that value when the circuit has been completely described. Suppose that a modification were introduced in the circuit by enlarging or diminishing the inclosed area by a small area lying wholly within a circle of radius $\frac{1}{2}\delta$. Consider the circuit $ABCDEA$ and the modified circuit $ABC'DEA$. As these circuits coincide except for the arcs BCD and $BC'D$, it is only necessary to show that f takes on the same value at D whether D is reached from B by the way of C or by the way of C'. But this is necessarily so for the reason that both arcs are within a circle of radius $\frac{1}{2}\delta$. Then the value of f must still return to its initial value $f(x, y)$ when the modified circuit is described. Now to complete the proof of the theorem, it suffices to note that any circuit which can be shrunk to nothing can be made up by piecing together a number of small circuits as shown in the figure. Then as the change in f around any one of the small circuits is zero, the change must be zero around 2, 3, 4, $\cdots$ adjacent circuits, and thus finally around the complete large circuit.



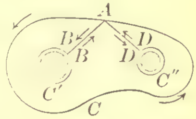


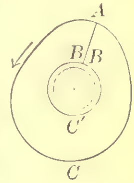
Reducibility of circuits. If a circuit can be shrunk away to nothing, it is said to be *reducible;* if it cannot, it is said to be *irreducible*. In a simply connected region all circuits are reducible; in a multiply connected region there are an infinity of irreducible circuits. Two circuits are said to be *equivalent* or reducible to each other when either can be expanded or shrunk into the other. The change in the value of f on passing around two equivalent circuits from A to A is the same, provided the circuits are described in the same direction. For consider the figure and the equivalent circuits ACA and $AC'A$ described as indicated by the large arrows. It is clear that either may be modified little by little, as indicated in the proof above, until it has been changed into the other. Hence the change in the value of f around the two circuits is the same. Or, as another proof, it may be observed that the combined circuit $ACAC'A$, where the second is described as indicated by the small arrows, may be regarded as a reducible circuit which touches itself at A. Then the change of f around the circuit is zero and f must lose as much on passing from A to A by C' as it gains in passing from A to A by C. Hence on passing from A to A by C' in the direction of the large arrows the gain in f must be the same as on passing by C.



It is now possible to see that *any circuit ABC may be reduced to circuits around the portions cut out of the region combined with lines going to and from A and the boundaries.* The figure shows this; for the circuit $ABC'BADC''DA$ is clearly

reducible to the circuit ACA. It must not be forgotten that although the lines AB and BA coincide, the values of the function are not necessarily the same on AB as on BA but differ by the amount of change introduced in f on passing around the irreducible circuit $BC''B$. One of the cases which arises most frequently in practice is that in which the successive branches of $f(x, y)$ differ by a constant amount as in the case $z = \tan^{-1}(y/x)$ where 2π is the difference between successive values of z for the same values of the variables. If now a circuit such as $ABC''BA$ be considered, where it is imagined that the origin lies within $BC''B$, it is clear that the values of z along AB and along BA differ by 2π, and whatever z gains on passing from A to B will be lost on passing from B to A, although the values through which z changes will be different in the two cases by the amount 2π. Hence the circuit $ABC''BA$ gives the same changes for z as the simpler circuit $BC''B$. In other words the result is obtained that *if the different values of a multiple valued function for the same values of the variables differ by a constant independent of the values of the variables, any circuit may be reduced to circuits about the boundaries of the portions removed;* in this case the lines going from the point A to the boundaries and back may be discarded.





EXERCISES

1. Draw the contour lines and sketch the surfaces corresponding to

$$(\alpha)\ z = \frac{x+y}{x-y}, \quad z(0, 0) = 0. \qquad (\beta)\ z = \frac{xy}{x+y}, \quad z(0, 0) = 0.$$

Note that here and in the text only one of the contour lines passes through the origin although an infinite number have it as a frontier point between two parts of the same contour line. Discuss the double limits $\lim\limits_{x \doteq 0} \lim\limits_{y \doteq 0} z$, $\lim\limits_{y \doteq 0} \lim\limits_{x \doteq 0} z$.

2. Draw the contour lines and sketch the surfaces corresponding to

$$(\alpha)\ z = \frac{x^2+y^2-1}{2y}, \qquad (\beta)\ z = \frac{y^2}{x}, \qquad (\gamma)\ z = \frac{x^2+2y^2-1}{2x^2+y^2-1}.$$

Examine particularly the behavior of the function in the neighborhood of the apparent points of intersection of different contour lines. Why apparent?

3. State and prove for functions of two independent variables the generalizations of Theorems 6–11 of Chap. II. Note that the theorem on uniformity is proved for two variables by the application of Ex. 9, p. 40, in almost the identical manner as for the case of one variable.

4. Outline definitions and theorems for functions of three variables. In particular indicate the contour surfaces of the functions

$$(\alpha)\ u = \frac{x+y+2z}{x-y-z}, \qquad (\beta)\ u = \frac{x^2+y^2+z^2}{x+y+z}, \qquad (\gamma)\ u = \frac{xy}{z},$$

and discuss the triple limits as x, y, z in different orders approach the origin.

5. Let $z = P(x, y)/Q(x, y)$, where P and Q are polynomials, be a rational function of x and y. Show that if the curves $P = 0$ and $Q = 0$ intersect in any points, all the contour lines of z will converge toward these points; and conversely show

that if two different contour lines of z apparently cut in some point, all the contour lines will converge toward that point, P and Q will there vanish, and z will be undefined.

6. If D is the minimum difference between different values of a multiple valued function, as in the text, and if the function returns to its initial value plus $D' \geqq D$ when P describes a circuit, show that it will return to its initial value plus $D' \geqq D$ when P describes the new circuit formed by piecing on to the given circuit a small region which lies within a circle of radius $\frac{1}{2}\delta$.

7. Study the function $z = \tan^{-1}(y/x)$, noting especially the relation between contour lines and the surface. To eliminate the origin at which the function is not defined draw a small circle about the point $(0, 0)$ and observe that the region of the whole xy-plane outside this circle is not simply connected but may be made so by drawing a cut from the circumference off to an infinite distance. Study the variation of the function as P describes various circuits.

8. Study the contour lines and the surfaces due to the functions

$$(\alpha)\ z = \tan^{-1} xy, \qquad (\beta)\ z = \tan^{-1}\frac{1 - x^2}{1 - y^2}, \qquad (\gamma)\ z = \sin^{-1}(x - y).$$

Cut out the points where the functions are not defined and follow the changes in the functions about such circuits as indicated in the figures of the text. How may the region of definition be made simply connected?

9. Consider the function $z = \tan^{-1}(P/Q)$ where P and Q are polynomials and where the curves $P = 0$ and $Q = 0$ intersect in n points $(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)$ but are not tangent (the polynomials have common solutions which are not multiple roots). Show that the value of the function will change by $2k\pi$ if (x, y) describes a circuit which includes k of the points. Illustrate by taking for P/Q the fractions in Ex. 2.

10. Consider regions or volumes in space. Show that there are regions in which some circuits cannot be shrunk away to nothing; also regions in which all circuits may be shrunk away but not all closed surfaces.

46. First partial derivatives. Let $z = f(x, y)$ be a single valued function, or one branch of a multiple valued function, defined for (a, b) and for all points in the neighborhood. If y be given the value b, then z becomes a function $f(x, b)$ of x alone, and if that function has a derivative for $x = a$, that derivative is called the *partial derivative* of $z = f(x, y)$ with respect to x at (a, b). Similarly, if x is held fast and equal to a and if $f(a, y)$ has a derivative when $y = b$, that derivative is called the partial derivative of z with respect to y at (a, b). To obtain these derivatives formally in the case of a given function $f(x, y)$ it is merely necessary to differentiate the function by the ordinary rules, treating y as a constant when finding the derivative with respect to x and x as a constant for the derivative with respect to y. Notations are

$$\frac{\partial f}{\partial x} = \frac{\partial z}{\partial x} = f'_x = f_x = z'_x = D_x f = D_x z = \left(\frac{dz}{dx}\right)_y$$

for the x-derivative with similar ones for the y-derivative. The partial derivatives are the limits of the quotients

$$\lim_{h \doteq 0} \frac{f(a+h,\, b) - f(a,\, b)}{h}, \qquad \lim_{k \doteq 0} \frac{f(a,\, b+k) - f(a,\, b)}{k}, \tag{2}$$

provided those limits exist. The application of the Theorem of the Mean to the functions $f(x, b)$ and $f(a, y)$ gives

$$\begin{aligned} f(a+h,\, b) - f(a,\, b) &= hf'_x(a+\theta_1 h,\, b), \quad 0 < \theta_1 < 1, \\ f(a,\, b+k) - f(a,\, b) &= kf'_y(a,\, b+\theta_2 k), \quad 0 < \theta_2 < 1, \end{aligned} \tag{3}$$

under the proper but evident restrictions (see § 26).

Two comments may be made. First, some writers denote the partial derivatives by the same symbols dz/dx and dz/dy as if z were a function of only one variable and were differentiated with respect to that variable; and if they desire especially to call attention to the other variables which are held constant, they affix them as subscripts as shown in the last symbol given (p. 93). This notation is particularly prevalent in thermodynamics. As a matter of fact, it would probably be impossible to devise a simple notation for partial derivatives which should be satisfactory for all purposes. The only safe rule to adopt is to use a notation which is sufficiently explicit for the purposes in hand, and at all times to pay careful attention to what the derivative actually means in each case. Second, it should be noted that for points on the boundary of the region of definition of $f(x, y)$ there may be merely right-hand or left-hand partial derivatives or perhaps none at all. For it is necessary that the lines $y = b$ and $x = a$ cut into the region on one side or the other in the neighborhood of (a, b) if there is to be a derivative even one-sided; and at a corner of the boundary it may happen that neither of these lines cuts into the region.

Theorem. If $f(x, y)$ and its derivatives f'_x and f'_y are continuous functions of (x, y) in the neighborhood of (a, b), the increment Δf may be written in any of the three forms

$$\begin{aligned} \Delta f &= f(a+h,\, b+k) - f(a,\, b) \\ &= hf'_x(a+\theta_1 h,\, b) + kf'_y(a+h,\, b+\theta_2 k) \\ &= hf'_x(a+\theta h,\, b+\theta k) + kf'_y(a+\theta h,\, b+\theta k) \\ &= hf'_x(a,\, b) + kf'_y(a,\, b) + \zeta_1 h + \zeta_2 k, \end{aligned} \tag{4}$$

where the θ's are proper fractions, the ζ's infinitesimals.

To prove the first form, add and subtract $f(a+h,\, b)$; then

$$\begin{aligned} \Delta f &= [f(a+h,\, b) - f(a,\, b)] + [f(a+h,\, b+k) - f(a+h,\, b)] \\ &= hf'_x(a+\theta_1 h,\, b) + kf'_y(a+h,\, b+\theta_2 k) \end{aligned}$$

by the application of the Theorem of the Mean for functions of a single variable (§§ 7, 26). The application may be made because the function is continuous and the indicated derivatives exist. Now if the derivatives are also continuous, they may be expressed as

$$f'_x(a+\theta_1 h,\, b) = f'_x(a,\, b) + \zeta_1, \qquad f'_y(a+h,\, b+\theta_2 k) = f'_y(a,\, b) + \zeta_2$$

where ζ_1, ζ_2 may be made as small as desired by taking h and k sufficiently small. Hence the third form follows from the first. The second form, which is symmetric in the increments h, k, may be obtained by writing $x = a + th$ and $y = b + tk$. Then $f(x, y) = \Phi(t)$. As f is continuous in (x, y), the function Φ is continuous in t and its increment is

$$\Delta\Phi = f(a + \overline{t + \Delta t}h, b + t + \Delta t k) - f(a + th, b + tk).$$

This may be regarded as the increment of f taken from the point (x, y) with $\Delta t \cdot h$ and $\Delta t \cdot k$ as increments in x and y. Hence $\Delta\Phi$ may be written as

$$\Delta\Phi = \Delta t \cdot hf'_x(a + th, b + tk) + \Delta t \cdot kf'_y(a + th, b + tk) + \zeta_1\Delta t \cdot h + \zeta_2\Delta t \cdot k.$$

Now if $\Delta\Phi$ be divided by Δt and Δt be allowed to approach zero, it is seen that

$$\lim \frac{\Delta\Phi}{\Delta t} = hf'_x(a + th, b + tk) + kf'_y(a + th, b + tk) = \frac{d\Phi}{dt}.$$

The Theorem of the Mean may now be applied to Φ to give $\Phi(1) - \Phi(0) = 1 \cdot \Phi'(\theta)$, and hence

$$\begin{aligned}\Phi(1) - \Phi(0) &= f(a + h, b + k) - f(a, b)\\ &= \Delta f = hf'_x(a + \theta h, b + \theta k) + kf'_y(a + \theta h, b + \theta k).\end{aligned}$$

47. The *partial differentials* of f may be defined as

$$\begin{aligned} d_x f &= f'_x\Delta x, \quad \text{so that} \quad dx = \Delta x, \quad \frac{d_x f}{dx} = \frac{\partial f}{\partial x},\\ d_y f &= f'_y\Delta y, \quad \text{so that} \quad dy = \Delta y, \quad \frac{d_y f}{dy} = \frac{\partial f}{\partial y}, \end{aligned} \tag{5}$$

where the indices x and y introduced in $d_x f$ and $d_y f$ indicate that x and y respectively are alone allowed to vary in forming the corresponding partial differentials. The *total differential*

$$df = d_x f + d_y f = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy, \tag{6}$$

which is the sum of the partial differentials, may be defined as that sum; but it is better defined as that part of the increment

$$\Delta f = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \zeta_1\Delta x + \zeta_2\Delta y \tag{7}$$

which is obtained by neglecting the terms $\zeta_1\Delta x + \zeta_2\Delta y$, which are of higher order than Δx and Δy. The total differential may therefore be computed by finding the partial derivatives, multiplying them respectively by dx and dy, and adding.

The total differential of $z = f(x, y)$ may be formed for (x_0, y_0) as

$$z - z_0 = \left(\frac{\partial f}{\partial x}\right)_0(x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0(y - y_0), \tag{8}$$

where the values $x - x_0$ and $y - y_0$ are given to the independent differentials dx and dy, and $df = dz$ is written as $z - z_0$. This, however, is

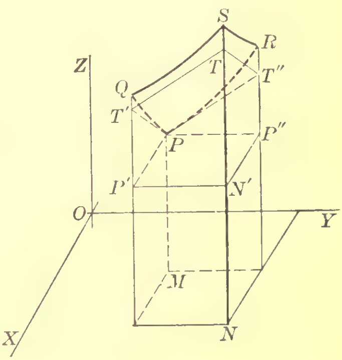
the equation of a plane since x and y are independent. The difference $\Delta f - df$ which measures the distance from the plane to the surface along a parallel to the z-axis is of higher order than $\sqrt{\Delta x^2 + \Delta y^2}$; for

$$\left|\frac{\Delta f - df}{\sqrt{\Delta x^2 + \Delta y^2}}\right| = \left|\frac{\zeta_1 \Delta x + \zeta_2 \Delta y}{\sqrt{\Delta x^2 + \Delta y^2}}\right| < |\zeta_1| + |\zeta_2| \doteq 0.$$

Hence the plane (8) will be defined as the *tangent plane* at (x_0, y_0, z_0) to the surface $z = f(x, y)$. The normal to the plane is

$$\frac{x - x_0}{\left(\frac{\partial f}{\partial x}\right)_0} = \frac{y - y_0}{\left(\frac{\partial f}{\partial y}\right)_0} = \frac{z - z_0}{-1}, \tag{9}$$

which will be defined as the *normal to the surface* at (x_0, y_0, z_0). The tangent plane will cut the planes $y = y_0$ and $x = x_0$ in lines of which the slope is f'_{x_0} and f'_{y_0}. The surface will cut these planes in curves which are tangent to the lines.



In the figure, $PQSR$ is a portion of the surface $z = f(x, y)$ and $PT'TT''$ is a corresponding portion of its tangent plane at $P(x_0, y_0, z_0)$. Now the various values may be read off.

$$PP' = \Delta x, \qquad P'Q = \Delta_x f,$$
$$P'T'/PP' = f'_x, \qquad P'T' = d_x f,$$
$$PP'' = \Delta y, \qquad P''R = \Delta_y f,$$
$$P''T''/PP'' = f'_y, \qquad P''T'' = d_y f,$$
$$P'T' + P''T'' = N'T, \qquad N'S = \Delta f,$$
$$N'T = df = d_x f + d_y f.$$

48. If the variables x and y are expressed as $x = \phi(t)$ and $y = \psi(t)$ so that $f(x, y)$ becomes a function of t, the derivative of f with respect to t is found from the expression for the increment of f.

$$\frac{\Delta f}{\Delta t} = \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \zeta_1 \frac{\Delta x}{\Delta t} + \zeta_2 \frac{\Delta y}{\Delta t}$$

or

$$\lim_{\Delta t \doteq 0} \frac{\Delta f}{\Delta t} = \frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}. \tag{10}$$

The conclusion requires that x and y should have finite derivatives with respect to t. The differential of f as a function of t is

$$df = \frac{df}{dt}dt = \frac{\partial f}{\partial x}\frac{dx}{dt}dt + \frac{\partial f}{\partial y}\frac{dy}{dt}dt = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \tag{11}$$

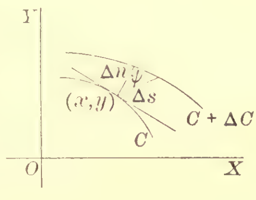
and hence it appears that *the differential has the same form as the total differential.* This result will be generalized later.

As a particular case of (10) suppose that x and y are so related that the point (x, y) moves along a line inclined at an angle τ to the x-axis. If s denote distance along the line, then

$$x = x_0 + s\cos\tau, \quad y = y_0 + s\sin\tau, \quad dx = \cos\tau ds, \quad dy = \sin\tau ds \qquad (12)$$

and
$$\frac{df}{ds} = \frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} = f'_x\cos\tau + f'_y\sin\tau. \qquad (13)$$

The derivative (13) is called the *directional derivative* of f in the direction of the line. The partial derivatives f'_x, f'_y are the particular directional derivatives along the directions of the x-axis and y-axis. The directional derivative of f in any direction is the rate of increase of f along that direction; if $z = f(x, y)$ be interpreted as a surface, the directional derivative is the slope of the curve in which a plane through the line (12) and perpendicular to the xy-plane cuts the surface. If $f(x, y)$ be represented by its contour lines, the derivative at a point (x, y) in any direction is the limit of the ratio $\Delta f/\Delta s = \Delta C/\Delta s$ of the increase of f, from one contour line to a neighboring one, to the distance between the lines in that direction. It is therefore evident that the derivative along any contour line is zero and that the derivative along the normal to the contour line is greater than in any other direction because the element dn of the normal is less than ds in any other direction. In fact, apart from infinitesimals of higher order,



$$\frac{\Delta n}{\Delta s} = \cos\psi, \quad \frac{\Delta f}{\Delta s} = \frac{\Delta f}{\Delta n}\cos\psi, \quad \frac{df}{ds} = \frac{df}{dn}\cos\psi. \qquad (14)$$

Hence it is seen that *the derivative along any direction may be found by multiplying the derivative along the normal by the cosine of the angle between that direction and the normal.* The derivative along the normal to a contour line is called the *normal derivative* of f and is, of course, a function of (x, y).

49. Next suppose that $u = f(x, y, z, \cdots)$ is a function of any number of variables. The reasoning of the foregoing paragraphs may be repeated without change except for the additional number of variables. The increment of f will take any of the forms

$$\begin{aligned} \Delta f &= f(a + h, b + k, c + l, \cdots) - f(a, b, c, \cdots) \\ &= hf'_x(a + \theta_1 h, b, c, \cdots) + kf'_y(a + h, b + \theta_2 k, c, \cdots) \\ &\quad + lf'_z(a + h, b + k, c + \theta_3 l, \cdots) + \cdots \\ &= [hf'_x + kf'_y + lf'_z + \cdots]_{a + \theta h, b + \theta k, c + \theta l, \cdots} \\ &= hf'_x + kf'_y + lf'_z + \cdots + \zeta_1 h + \zeta_2 k + \zeta_3 l + \cdots. \end{aligned}$$

and the total differential will naturally be defined as

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz + \cdots, \tag{16}$$

and finally if $x, y, z, \cdots$ be functions of t, it follows that

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} + \cdots \tag{17}$$

and the differential of f as a function of t is still (16).

If the variables $x, y, z, \cdots$ were expressed in terms of several new variables $r, s, \cdots$, the function f would become a function of those variables. To find the partial derivative of f with respect to one of those variables, say r, the remaining ones, $s, \cdots$, would be held constant and f would for the moment become a function of r alone, and so would $x, y, z, \cdots$. Hence (17) may be applied to obtain the partial derivatives

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial r} + \cdots,$$

and

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial s} + \cdots, \text{ etc.} \tag{18}$$

These are the formulas for *change of variable* analogous to (4) of § 2. If these equations be multiplied by $\Delta r, \Delta s, \cdots$ and added,

$$\frac{\partial f}{\partial r}\Delta r + \frac{\partial f}{\partial s}\Delta s + \cdots = \frac{\partial f}{\partial x}\left(\frac{\partial x}{\partial r}\Delta r + \frac{\partial x}{\partial s}\Delta s + \cdots\right) + \frac{\partial f}{\partial y}\left(\frac{\partial y}{\partial r}\Delta r + \cdots\right) + \cdots,$$

or

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz + \cdots;$$

for when $r, s, \cdots$ are the independent variables, the parentheses above are $dx, dy, dz, \cdots$ and the expression on the left is df.

Theorem. The expression of the total differential of a function of $x, y, z, \cdots$ as $df = f'_x dx + f'_y dy + f'_z dz + \cdots$ is the same whether $x, y, z, \cdots$ are the independent variables or functions of other independent variables $r, s, \cdots$; it being assumed that all the derivatives which occur, whether of f by $x, y, z, \cdots$ or of $x, y, z, \cdots$ by $r, s, \cdots$, are continuous functions.

By the same reasoning or by virtue of this theorem the rules

$$d(cu) = c\,du, \quad d(u + v - w) = du + dv - dw,$$
$$d(uv) = u\,dv + v\,du, \quad d\left(\frac{u}{v}\right) = \frac{v\,du - u\,dv}{v^2}, \tag{19}$$

of the differential calculus will apply to calculate the total differential of combinations or functions of several variables. If by this means, or any other, there is obtained an expression

$$df = R(r, s, t, \cdots)\,dr + S(r, s, t, \cdots)\,ds + T(r, s, t, \cdots)\,dt + \cdots \tag{20}$$

for the total differential in which $r, s, t, \cdots$ are *independent* variables, the coefficients $R, S, T, \cdots$ are the derivatives

$$R = \frac{\partial f}{\partial r}, \quad S = \frac{\partial f}{\partial s}, \quad T = \frac{\partial f}{\partial t}, \cdots. \tag{21}$$

For in the equation $df = Rdr + Sds + Tdt + \cdots = f_r'dr + f_s'ds + f_t'dt + \cdots$, the variables $r, s, t, \cdots$, being independent, may be assigned increments absolutely at pleasure and if the particular choice $dr = 1$, $ds = dt = \cdots = 0$, be made, it follows that $R = f_r'$; and so on. The single equation (20) is thus equivalent to the equations (21) in number equal to the number of the independent variables.

As an example, consider the case of the function $\tan^{-1}(y/x)$. By the rules (19),

$$d\tan^{-1}\frac{y}{x} = \frac{d(y/x)}{1 + (y/x)^2} = \frac{dy/x - ydx/x^2}{1 + (y/x)^2} = \frac{xdy - ydx}{x^2 + y^2}.$$

Then $$\frac{\partial}{\partial x}\tan^{-1}\frac{y}{x} = -\frac{y}{x^2 + y^2}, \qquad \frac{\partial}{\partial y}\tan^{-1}\frac{y}{x} = \frac{x}{x^2 + y^2}, \qquad \text{by (20)–(21).}$$

If y and x were expressed as $y = \sinh rst$ and $x = \cosh rst$, then

$$d\tan^{-1}\frac{y}{x} = \frac{xdy - ydx}{x^2 + y^2} = \frac{[stdr + rtds + rsdt][\cosh^2 rst - \sinh^2 rst]}{\cosh^2 rst + \sinh^2 rst}$$

and $$\frac{\partial f}{\partial r} = \frac{st}{\cosh 2\,rst}, \qquad \frac{\partial f}{\partial s} = \frac{rt}{\cosh 2\,rst}, \qquad \frac{\partial f}{\partial t} = \frac{rs}{\cosh 2\,rst}.$$

EXERCISES

1. Find the partial derivatives f_x', f_y' or f_x', f_y', f_z' of these functions:

(α) $\log(x^2 + y^2)$, (β) $e^x \cos y \sin z$, (γ) $x^2 + 3xy + y^3$,

(δ) $\dfrac{xy}{x + y}$, (ϵ) $\dfrac{e^{xy}}{e^x + e^y}$, ($\zeta$) $\log(\sin x + \sin^2 y + \sin^3 z)$,

(η) $\sin^{-1}\dfrac{y}{x}$, (θ) $\dfrac{z}{x}e^{\frac{y}{x}}$, ($\iota$) $\tanh^{-1}\sqrt{2}\left(\dfrac{xy + yz + zx}{x^2 + y^2 + z^2}\right)^{\frac{1}{2}}$.

2. Apply the definition (2) directly to the following to find the partial derivatives at the indicated points:

(α) $\dfrac{xy}{x + y}$ at $(1, 1)$, (β) $x^2 + 3xy + y^3$ at $(0, 0)$, and (γ) at $(1, 1)$,

(δ) $\dfrac{x - y}{x + y}$ at $(0, 0)$; also try differentiating and substituting $(0, 0)$.

3. Find the partial derivatives and hence the total differential of:

(α) $\dfrac{e^{xy}}{x^2 + y^2}$, ($\beta$) $x \log yz$, (γ) $\sqrt{a^2 - x^2 - y^2}$,

(δ) $e^{-x}\sin y$, (ϵ) $e^{z^2}\sinh xy$, (ζ) $\log\tan\left(x + \dfrac{\pi}{4}y\right)$,

(η) $\left(\dfrac{y}{z}\right)^x$, ($\theta$) $\dfrac{x - y}{x + z}$, (ι) $\log\left(\dfrac{3x}{y^2} + \sqrt{1 + \dfrac{z^2x^2}{y^4}}\right)$.

4. Find the general equations of the tangent plane and normal line to these surfaces and find the equations of the plane and line for the indicated (x_0, y_0):

(α) the helicoid $z = k\tan^{-1}(y/x)$, $(1, 0), (1, -1), (0, 1)$,
(β) the paraboloid $4pz = (x^2 + y^2)$, $(0, p), (2p, 0), (p, -p)$,
(γ) the hemisphere $z = \sqrt{a^2 - x^2 - y^2}$, $(0, -\frac{1}{2}a), (\frac{1}{2}a, \frac{1}{2}a), (\frac{1}{2}\sqrt{3}a, 0)$,
(δ) the cubic $xyz = 1$, $(1, 1, 1), (-\frac{1}{2}, -\frac{1}{2}, 4), (4, \frac{1}{2}, \frac{1}{2})$.

5. Find the derivative with respect to t in these cases by (10):

(α) $f = x^2 + y^2$, $x = a\cos t$, $y = b\sin t$, (β) $\tan^{-1}\sqrt{\frac{y}{x}}$, $y = \cosh t$, $x = \sinh t$,

(γ) $\sin^{-1}(x - y)$, $x = 3t$, $y = 4t^3$, (δ) $\cos 2xy$, $x = \tan^{-1}t$, $y = \cot^{-1}t$.

6. Find the directional derivative in the direction indicated and obtain its numerical value at the points indicated:

(α) x^2y, $\tau = 45°$, $(1, 2)$, (β) $\sin^2 xy$, $\tau = 60°$, $(\sqrt{3}, -2)$.

7. (α) Determine the maximum value of df/ds from (13) by regarding τ as variable and applying the ordinary rules. Show that the direction that gives the maximum is

$$\tau = \tan^{-1}\frac{f_y'}{f_x'}, \quad \text{and then} \quad \frac{df}{dn} = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}.$$

(β) Show that the sum of the squares of the derivatives along any two perpendicular directions is the same and is the square of the normal derivative.

8. Show that $(f_x' + y'f_y')/\sqrt{1 + y'^2}$ and $(f_x'y' - f_y')/\sqrt{1 + y'^2}$ are the derivatives of f along the curve $y = \phi(x)$ and normal to the curve.

9. If df/dn is defined by the work of Ex. 7 (α), prove (14) as a consequence.

10. Apply the formulas for the change of variable to the following cases:

(α) $r = \sqrt{x^2 + y^2}$, $\phi = \tan^{-1}\frac{y}{x}$. Find $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$.

(β) $x = r\cos\phi$, $y = r\sin\phi$. Find $\frac{\partial f}{\partial r}, \frac{\partial f}{\partial \phi}, \left(\frac{\partial f}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial f}{\partial \phi}\right)^2$.

(γ) $x = 2r - 3s + 7$, $y = -r + 8s - 9$. Find $\frac{\partial u}{\partial r} = 4x + 2y$ if $u = x^2 - y^2$.

(δ) $\begin{cases} x = x'\cos\alpha - y'\sin\alpha, \\ y = x'\sin\alpha + y'\cos\alpha. \end{cases}$ Show $\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 = \left(\frac{\partial f}{\partial x'}\right)^2 + \left(\frac{\partial f}{\partial y'}\right)^2$.

(ϵ) Prove $\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} = 0$ if $f(u, v) = f(x - y, y - x)$.

(ζ) Let $x = ax' + by' + cz'$, $y = a'x' + b'y' + c'z'$, $z = a''x' + b''y' + c''z'$, where $a, b, c, a', b', c', a'', b'', c''$ are the direction cosines of new rectangular axes with respect to the old. This transformation is called an *orthogonal transformation*. Show

$$\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial f}{\partial z}\right)^2 = \left(\frac{\partial f}{\partial x'}\right)^2 + \left(\frac{\partial f}{\partial y'}\right)^2 + \left(\frac{\partial f}{\partial z'}\right)^2 = \left(\frac{df}{dn}\right)^2.$$

11. Define directional derivative in space; also normal derivative and establish (14) for this case. Find the normal derivative of $f = xyz$ at $(1, 2, 3)$.

12. Find the total differential and hence the partial derivatives in Exs. 1, 3, and

(α) $\log(x^2 + y^2 + z^2)$, (β) y/x, (γ) $x^2ye^{xy^2}$, (δ) $xyz\log xyz$,

(ϵ) $u = x^2 - y^2$, $x = r\cos st$, $y = s\sin rt$. Find $\partial u/\partial r$, $\partial u/\partial s$, $\partial u/\partial t$.

(ζ) $u = y/x$, $x = r\cos\phi\sin\theta$, $y = r\sin\phi\sin\theta$. Find u'_r, u'_ϕ, u'_θ.

(η) $u = e^{xy}$, $x = \log\sqrt{r^2+s^2}$, $y = \tan^{-1}(s/r)$. Find u'_r, u'_s.

13. If $\dfrac{\partial f}{\partial x} = \dfrac{\partial g}{\partial y}$ and $\dfrac{\partial f}{\partial y} = -\dfrac{\partial g}{\partial x}$, show $\dfrac{\partial f}{\partial r} = \dfrac{1}{r}\dfrac{\partial g}{\partial \phi}$ and $\dfrac{1}{r}\dfrac{\partial f}{\partial \phi} = -\dfrac{\partial g}{\partial r}$ if r, ϕ are polar coördinates and f, g are any two functions.

14. If $p(x, y, z, t)$ is the pressure in a fluid, or $\rho(x, y, z, t)$ is the density, depending on the position in the fluid and on the time, and if u, v, w are the velocities of the particles of the fluid along the axes,

$$\frac{dp}{dt} = u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y} + w\frac{\partial p}{\partial z} + \frac{\partial p}{\partial t} \quad\text{and}\quad \frac{d\rho}{dt} = u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z} + \frac{\partial \rho}{dt}.$$

Explain the meaning of each derivative and prove the formula.

15. If $z = xy$, interpret z as the area of a rectangle and mark d_xz, Δ_yz, Δz on the figure. Consider likewise $u = xyz$ as the volume of a rectangular parallelepiped.

16. *Small errors.* If $f(x, y)$ be a quantity determined by measurements on x and y, the error in f due to small errors dx, dy in x and y may be estimated as $df = f'_x dx + f'_y dy$ and the relative error may be taken as $df \div f = d\log f$. Why is this?

(α) Suppose $S = \frac{1}{2}ab\sin C$ be the area of a triangle with $a = 10$, $b = 20$, $C = 30°$. Find the error and the relative error if a is subject to an error of 0.1. *Ans.* 0.5, 1%.

(β) In (α) suppose C were liable to an error of 10′ of arc. *Ans.* 0.27, $\frac{1}{2}$%.

(γ) If a, b, C are liable to errors of 1%, the combined error in S may be 3.1%.

(δ) The radius r of a capillary tube is determined from $13.6\,\pi r^2 l = w$ by finding the weight w of a column of mercury of length l. If $w = 1$ gram with an error of 10^{-3} gr. and $l = 10$ cm. with an error of 0.2 cm., determine the possible error and relative error in r. *Ans.* 1.05%, 5×10^{-4}, mostly due to error in l.

(ϵ) The formula $c^2 = a^2 + b^2 - 2ab\cos C$ is used to determine c where $a = 20$, $b = 20$, $C = 60°$ with possible errors of 0.1 in a and b and 30′ in C. Find the possible absolute and relative errors in c. *Ans.* $\frac{1}{4}$, $1\frac{1}{4}$%.

(ζ) The possible percentage error of a product is the sum of the percentage errors of the factors.

(η) The constant g of gravity is determined from $g = 2st^{-2}$ by observing a body fall. If s is set at 4 ft. and t determined at about $\frac{1}{2}$ sec., show that the error in g is almost wholly due to the error in t, that is, that s can be set very much more accurately than t can be determined. For example, find the error in t which would make the same error in g as an error of $\frac{1}{8}$ inch in s.

(θ) The constant g is determined by $gt^2 = \pi^2 l$ with a pendulum of length l and period t. Suppose t is determined by taking the time 100 sec. of 100 beats of the pendulum with a stop watch that measures to $\frac{1}{5}$ sec. and that l may be measured as 100 cm. accurate to $\frac{1}{2}$ millimeter. Discuss the errors in g.

17. Let the coördinate x of a particle be $x = f(q_1, q_2)$ and depend on two independent variables q_1, q_2. Show that the velocity and kinetic energy are

$$v = f'_{q_1}\frac{dq_1}{dt} + f'_{q_2}\frac{dq_2}{dt}, \qquad T = \tfrac{1}{2}mv^2 = a_{11}\dot{q}_1^2 + 2a_{12}\dot{q}_1\dot{q}_2 + a_{22}\dot{q}_2^2,$$

where dots denote differentiation by t, and a_{11}, a_{12}, a_{22} are functions of (q_1, q_2). Show $\dfrac{\partial v}{\partial q_i} = \dfrac{\partial x}{\partial q_i}$, $i = 1, 2$, and similarly for any number of variables q.

18. The helix $x = a\cos t$, $y = a\sin t$, $z = at\tan\alpha$ cuts the sphere $x^2 + y^2 + z^2 = a^2\sec^2\beta$ at $\sin^{-1}(\sin\alpha\sin\beta)$.

19. Apply the Theorem of the Mean to prove that $f(x, y, z)$ is a constant if $f'_x = f'_y = f'_z = 0$ is true for all values of x, y, z. Compare Theorem 16 (§ 27) and make the statement accurate.

20. Transform $\dfrac{df}{dn} = \sqrt{\left(\dfrac{\partial f}{\partial x}\right)^2 + \left(\dfrac{\partial f}{\partial y}\right)^2 + \left(\dfrac{\partial f}{\partial z}\right)^2}$ to (α) cylindrical and (β) polar coördinates (§ 40).

21. Find the angle of intersection of the helix $x = 2\cos t$, $y = 2\sin t$, $z = t$ and the surface $xyz = 1$ at their first intersection, that is, with $0 < t < \frac{1}{4}\pi$.

22. Let f, g, h be three functions of (x, y, z). In cylindrical coördinates (§ 40) form the combinations $F = f\cos\phi + g\sin\phi$, $G = -f\sin\phi + g\cos\phi$, $H = h$. Transform

$$(\alpha)\ \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z}, \qquad (\beta)\ \frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}, \qquad (\gamma)\ \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}$$

to cylindrical coördinates and express in terms of F, G, H in simplest form.

23. Given the functions y^x and $(z^y)^x$ and $z^{(y^x)}$. Find the total differentials and hence obtain the derivatives of x^x and $(x^x)^x$ and $x^{(x^x)}$.

50. Derivatives of higher order. If the first derivatives be again differentiated, there arise four derivatives f''_{xx}, f''_{xy}, f''_{yx}, f''_{yy} of the second order, where the first subscript denotes the first differentiation. These may also be written

$$f''_{xx} = \frac{\partial^2 f}{\partial x^2}, \quad f''_{xy} = \frac{\partial^2 f}{\partial y \partial x}, \quad f''_{yx} = \frac{\partial^2 f}{\partial x \partial y}, \quad f''_{yy} = \frac{\partial^2 f}{\partial y^2},$$

where the derivative of $\partial f/\partial y$ with respect to x is written $\partial^2 f/\partial x \partial y$ with the variables in the same order as required in $D_x D_y f$ and opposite to the order of the subscripts in f''_{yx}. This matter of order is usually of no importance owing to the theorem: *If the derivatives f'_x, f'_y have derivatives f''_{xy}, f''_{yx} which are continuous in (x, y) in the neighborhood of any point (x_0, y_0), the derivatives f''_{xy} and f''_{yx} are equal*, that is, $f''_{xy}(x_0, y_0) = f''_{yx}(x_0, y_0)$.

The theorem may be proved by repeated application of the Theorem of the Mean. For

$$[f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)] - [f(x_0 + h, y_0) - f(x_0, y_0)] = [\phi(y_0 + k) - \phi(y_0)]$$
$$= [f(x_0 + h, y_0 + k) - f(x_0 + h, y_0)] - [f(x_0, y_0 + k) - f(x_0, y_0)] = [\psi(x_0 + h) - \psi(x_0)]$$

where $\phi(y)$ stands for $f(x_0 + h, y) - f(x_0, y)$ and $\psi(x)$ for $f(x, y_0 + k) - f(x, y_0)$. Now

$$\phi(y_0 + k) - \phi(y_0) = k\phi'(y_0 + \theta k) = k[f'_y(x_0 + h, y_0 + \theta k) - f'_y(x_0, y_0 + \theta k)],$$
$$\psi(x_0 + h) - \psi(x_0) = h\psi'(x_0 + \theta' h) = h[f'_x(x_0 + \theta' h, y_0 + k) - f'_x(x_0 + \theta' h, y_0)]$$

by applying the Theorem of the Mean to $\phi(y)$ and $\psi(x)$ regarded as functions of a single variable and then substituting. The results obtained are necessarily equal to each other; but each of these is in form for another application of the theorem.

$$k[f'_y(x_0 + h, y_0 + \theta k) - f'_y(x_0, y_0 + \theta k)] = khf''_{yx}(x_0 + \eta h, y_0 + \theta k),$$
$$h[f'_x(x_0 + \theta' h, y_0 + k) - f'_x(x_0 + \theta' h, y_0)] = hkf''_{xy}(x_0 + \theta' h, y_0 + \eta' k).$$

Hence $$f''_{yx}(x_0 + \eta h, y_0 + \theta k) = f''_{xy}(x_0 + \theta' h, y_0 + \eta' k).$$

As the derivatives f''_{yx}, f''_{xy} are supposed to exist and be continuous in the variables (x, y) at and in the neighborhood of (x_0, y_0), the limit of each side of the equation exists as $h \doteq 0$, $k \doteq 0$ and the equation is true in the limit. Hence

$$f''_{yx}(x_0, y_0) = f''_{xy}(x_0, y_0).$$

The differentiation of the three derivatives f''_{x^2}, $f''_{xy} = f''_{yx}$, f''_{y^2} will give six derivatives of the third order. Consider f'''_{xxy} and f'''_{xyx}. These may be written as $(f'_x)''_{xy}$ and $(f'_x)''_{yx}$ and are equal by the theorem just proved (provided the restrictions as to continuity and existence are satisfied). A similar conclusion holds for f'''_{yxy} and f'''_{yyx}; the number of distinct derivatives of the third order reduces from six to four, just as the number of the second order reduces from four to three. In like manner for derivatives of any order, *the value of the derivative depends not on the order in which the individual differentiations with respect to x and y are performed, but only on the total number of differentiations with respect to each*, and the result may be written with the differentiations collected as

$$D_x^m D_y^n f = \frac{\partial^{m+n} f}{\partial x^m \partial y^n} = f^{(m+n)}_{x^m y^n}, \text{ etc.} \tag{22}$$

Analogous results hold for functions of any number of variables. If several derivatives are to be found and added together, a symbolic form of writing is frequently advantageous. For example,

$$(D_x^2 D_y D_z^3 + D_y^6)f = \frac{\partial^6 f}{\partial x^2 \partial y \partial z^3} + \frac{\partial^6 f}{\partial y^6}$$

or $$(D_x + D_y)^2 f = (D_x^2 + 2D_x D_y + D_y^2)f = f''_{x^2} + 2f''_{xy} + f''_{y^2}.$$

51. It is sometimes necessary to *change the variable* in higher derivatives, particularly in those of the second order. This is done by a repeated application of (18). Thus f''_{r^2} would be found by differentiating the first equation with respect to r, and f''_{rs} by differentiating the first by s or the second by r, and so on. Compare p. 12. The exercise below illustrates the method. It may be remarked that the use of *higher differentials* is often of advantage, although these differentials, like the higher differentials of functions of a single variable (Exs. 10, 16–19, p. 67), have the disadvantage that their form depends on what the independent variables are. This is also illustrated below. It should be particularly borne in mind that the great value of the first differential

lies in the facts that it may be treated like a finite quantity and that its form is independent of the variables.

To change the variable in $v''_{xx} + v''_{yy}$ to polar coördinates and show

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = \frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} + \frac{1}{r^2}\frac{\partial^2 v}{\partial \phi^2}, \qquad \begin{cases} x = r\cos\phi, & y = r\sin\phi, \\ r = \sqrt{x^2+y^2}, & \phi = \tan^{-1}(y/x). \end{cases}$$

Then
$$\frac{\partial v}{\partial x} = \frac{\partial v}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial v}{\partial \phi}\frac{\partial \phi}{\partial x}, \qquad \frac{\partial v}{\partial y} = \frac{\partial v}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial v}{\partial \phi}\frac{\partial \phi}{\partial y}$$

by applying (18) directly with x, y taking the place of r, s, $\cdots$ and r, ϕ the place of x, y, z, $\cdots$. These expressions may be reduced so that

$$\frac{\partial v}{\partial x} = \frac{\partial v}{\partial r}\frac{x}{\sqrt{x^2+y^2}} + \frac{\partial v}{\partial \phi}\frac{-y}{x^2+y^2} = \frac{\partial v}{\partial r}\frac{x}{r} + \frac{\partial v}{\partial \phi}\frac{-y}{r^2}.$$

Next
$$\frac{\partial^2 v}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial v}{\partial x} = \frac{\partial}{\partial r}\frac{\partial v}{\partial x}\cdot\frac{\partial r}{\partial x} + \frac{\partial}{\partial \phi}\frac{\partial v}{\partial x}\cdot\frac{\partial \phi}{\partial x}$$
$$= \left[\frac{\partial^2 v}{\partial r^2}\frac{x}{r} + \frac{\partial v}{\partial r}\frac{\partial}{\partial r}\frac{x}{r} + \frac{\partial^2 v}{\partial r\partial \phi}\frac{-y}{r^2} + \frac{\partial v}{\partial \phi}\frac{\partial}{\partial r}\frac{-y}{r^2}\right]\frac{x}{r}$$
$$+ \left[\frac{\partial^2 v}{\partial \phi\partial r}\frac{x}{r} + \frac{\partial v}{\partial r}\frac{\partial}{\partial \phi}\frac{x}{r} + \frac{\partial^2 v}{\partial \phi^2}\frac{-y}{r^2} + \frac{\partial v}{\partial \phi}\frac{\partial}{\partial \phi}\frac{-y}{r^2}\right]\frac{-y}{r^2}.$$

The differentiations of x/r and $-y/r^2$ may be performed as indicated with respect to r, ϕ, remembering that, as r, ϕ are independent, the derivative of r by ϕ is 0. Then

$$\frac{\partial^2 v}{\partial x^2} = \frac{x^2}{r^2}\frac{\partial^2 v}{\partial r^2} + \frac{y^2}{r^3}\frac{\partial v}{\partial r} - 2\frac{xy}{r^3}\frac{\partial^2 v}{\partial r\partial \phi} + 2\frac{xy}{r^4}\frac{\partial v}{\partial \phi} + \frac{y^2}{r^4}\frac{\partial^2 v}{\partial \phi^2}.$$

In like manner $\partial^2 v/\partial y^2$ may be found, and the sum of the two derivatives reduces to the desired expression. This method is long and tedious though straightforward.

It is considerably shorter to start with the expression in polar coördinates and transform by the same method to the one in rectangular coördinates. Thus

$$\frac{\partial v}{\partial r} = \frac{\partial v}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial v}{\partial y}\frac{\partial y}{\partial r} = \frac{\partial v}{\partial x}\cos\phi + \frac{\partial v}{\partial y}\sin\phi = \frac{1}{r}\left(\frac{\partial v}{\partial x}x + \frac{\partial v}{\partial y}y\right),$$
$$\frac{\partial}{\partial r}\left(r\frac{\partial v}{\partial r}\right) = \left(\frac{\partial^2 v}{\partial x^2}\cos\phi + \frac{\partial^2 v}{\partial y\partial x}\sin\phi\right)x + \left(\frac{\partial^2 v}{\partial x\partial y}\cos\phi + \frac{\partial^2 v}{\partial y^2}\sin\phi\right)y + \frac{\partial v}{\partial x}\cos\phi + \frac{\partial v}{\partial y}\sin\phi,$$
$$\frac{\partial v}{\partial \phi} = \frac{\partial v}{\partial x}\frac{\partial x}{\partial \phi} + \frac{\partial v}{\partial y}\frac{\partial y}{\partial \phi} = -\frac{\partial v}{\partial x}r\sin\phi + \frac{\partial v}{\partial y}r\cos\phi = -\frac{\partial v}{\partial x}y + \frac{\partial v}{\partial y}x,$$
$$\frac{1}{r}\frac{\partial^2 v}{\partial \phi^2} = \left(\frac{\partial^2 v}{\partial x^2}\sin\phi - \frac{\partial^2 v}{\partial y\partial x}\cos\phi\right)y + \left(-\frac{\partial^2 v}{\partial x\partial y}\sin\phi + \frac{\partial^2 v}{\partial y^2}\cos\phi\right)x$$
$$- \frac{\partial v}{\partial x}\cos\phi - \frac{\partial v}{\partial y}\sin\phi.$$

Then
$$\frac{\partial}{\partial r}\left(r\frac{\partial v}{\partial r}\right) + \frac{1}{r}\frac{\partial^2 v}{\partial \phi^2} = \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right)r$$

or
$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial v}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 v}{\partial \phi^2} = \frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} + \frac{1}{r^2}\frac{\partial^2 v}{\partial \phi^2}. \tag{23}$$

The definitions $d_x^2 f = f''_{xx}dx^2$, $d_x d_y f = f''_{xy}dxdy$, $d_y^2 f = f''_{yy}dy^2$ would naturally be given for *partial differentials of the second order*, each of which would vanish if f reduced to either of the independent variables x, y or to any linear function of them. Thus the second differentials of the independent variables are zero. The

second total differential would be obtained by differentiating the first total differential.

$$d^2f = ddf = d\left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy\right) = d\frac{\partial f}{\partial x}dx + d\frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial x}d^2x + \frac{\partial f}{\partial y}d^2y;$$

but $$d\frac{\partial f}{\partial x} = \frac{\partial^2 f}{\partial x^2}dx + \frac{\partial^2 f}{\partial y\partial x}dy, \qquad d\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial x\partial y}dx + \frac{\partial^2 f}{\partial y^2}dy,$$

and $$d^2f = \frac{\partial^2 f}{\partial x^2}dx^2 + 2\frac{\partial^2 f}{\partial x\partial y}dxdy + \frac{\partial^2 f}{\partial y^2}dy^2 + \frac{\partial f}{\partial x}d^2x + \frac{\partial f}{\partial y}d^2y. \tag{24}$$

The last two terms vanish and the total differential reduces to the first three terms if x and y are the independent variables; and in this case the second derivatives, $f''_{xx}, f''_{xy}, f''_{yy}$, are the coefficients of dx^2, $2\,dxdy$, dy^2, which enables those derivatives to be found by an extension of the method of finding the first derivatives (§ 49). The method is particularly useful when all the second derivatives are needed.

The problem of the change of variable may now be treated. Let

$$\begin{aligned} d^2v &= \frac{\partial^2 v}{\partial x^2}dx^2 + 2\frac{\partial^2 v}{\partial x^2}dxdy + \frac{\partial^2 v}{\partial y^2}dy^2 \\ &= \frac{\partial^2 v}{\partial r^2}dr^2 + 2\frac{\partial^2 v}{\partial r\partial\phi}drd\phi + \frac{\partial^2 v}{\partial\phi^2}d\phi^2 + \frac{\partial v}{\partial r}d^2r + \frac{\partial v}{\partial\phi}d^2\phi, \end{aligned}$$

where x, y are the independent variables and r, ϕ other variables dependent on them — in this case, defined by the relations for polar coördinates. Then

$$dx = \cos\phi dr - r\sin\phi d\phi, \qquad dy = \sin\phi dr + r\cos\phi d\phi$$

or $$dr = \cos\phi dx + \sin\phi dy, \qquad rd\phi = -\sin\phi dx + \cos\phi dy. \tag{25}$$

Then $$d^2r = (-\sin\phi dx + \cos\phi dy)\,d\phi = rd\phi d\phi = rd\phi^2.$$

$$drd\phi + rd^2\phi = -(\cos\phi dx + \sin\phi dy)\,d\phi = -drd\phi,$$

where the differentials of dr and $rd\phi$ have been found subject to $d^2x = d^2y = 0$. Hence $d^2r = rd\phi^2$ and $rd^2\phi = -2\,drd\phi$. These may be substituted in d^2v which becomes

$$d^2v = \frac{\partial^2 v}{\partial r^2}dr^2 + 2\left(\frac{\partial^2 v}{\partial r\partial\phi} - \frac{1}{r}\frac{\partial v}{\partial\phi}\right)drd\phi + \left(\frac{\partial^2 v}{\partial\phi^2} + r\frac{\partial v}{\partial r}\right)d\phi^2.$$

Next the values of dr^2, $drd\phi$, $d\phi^2$ may be substituted from (25) and

$$\begin{aligned} d^2v &= \left[\frac{\partial^2 v}{\partial r^2}\cos^2\phi - \frac{2}{r}\left(\frac{\partial^2 v}{\partial r\partial\phi} - \frac{1}{r}\frac{\partial v}{\partial\phi}\right)\cos\phi\sin\phi + \left(\frac{\partial^2 v}{\partial\phi^2} + r\frac{\partial v}{\partial r}\right)\frac{\sin^2\phi}{r^2}\right]dx^2 \\ &+ 2\left[\frac{\partial^2 v}{\partial r^2}\cos\phi\sin\phi + \left(\frac{\partial^2 v}{\partial r\partial\phi} - \frac{1}{r}\frac{\partial v}{\partial\phi}\right)\frac{\cos^2\phi - \sin^2\phi}{r} - \frac{\partial^2 v}{\partial\phi^2}\frac{\cos\phi\sin\phi}{r^2}\right]dxdy \\ &+ \left[\frac{\partial^2 v}{\partial r^2}\sin^2\phi + \frac{2}{r}\left(\frac{\partial^2 v}{\partial r\partial\phi} - \frac{1}{r}\frac{\partial v}{\partial\phi}\right)\cos\phi\sin\phi + \left(\frac{\partial^2 v}{\partial\phi^2} + r\frac{\partial v}{\partial r}\right)\frac{\cos^2\phi}{r^2}\right]dy^2. \end{aligned}$$

Thus finally the derivatives v''_{xx}, v''_{xy}, v''_{yy} are the three brackets which are the coefficients of dx^2, $2\,dxdy$, dy^2. The value of $v''_{xx} + v''_{yy}$ is as found before.

52. The condition $f''_{xy} = f''_{yx}$ which subsists in accordance with the fundamental theorem of § 50 gives *the condition that*

$$M(x, y)dx + N(x, y)dy = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = df$$

be the total differential of some function $f(x, y)$. *In fact*

$$\frac{\partial}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial M}{\partial y} = \frac{\partial N}{\partial x} = \frac{\partial}{\partial x}\frac{\partial f}{\partial y}$$

and
$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x} \quad \text{or} \quad \left(\frac{dM}{dy}\right)_x = \left(\frac{dN}{dx}\right)_y. \tag{26}$$

The second form, where the variables which are constant during the differentiation are explicitly indicated as subscripts, is more common in works on thermodynamics. It will be proved later that conversely if this relation (26) holds, the expression $Mdx + Ndy$ is the total differential of some function, and the method of finding the function will also be given (§§ 92, 124). In case $Mdx + Ndy$ is the differential of some function $f(x, y)$ it is usually called an *exact differential.*

The application of the condition for an exact differential may be made in connection with a problem in thermodynamics. Let S and U be the entropy and energy of a gas or vapor inclosed in a receptacle of volume v and subjected to the pressure p at the temperature T. The fundamental equation of thermodynamics, connecting the differentials of energy, entropy, and volume, is

$$dU = TdS - pdv; \quad \text{and} \quad \left(\frac{dT}{dv}\right)_S = -\left(\frac{dp}{dS}\right)_v \tag{27}$$

is the condition that dU be a total differential. Now, any two of the five quantities U, S, v, T, p may be taken as independent variables. In (27) the choice is S, v; if the equation were solved for dS, the choice would be U, v; and U, S if solved for dv. In each case the cross differentiation to express the condition (26) would give rise to a relation between the derivatives.

If p, T were desired as independent variables, the change of variable

$$dS = \left(\frac{dS}{dp}\right)_T dp + \left(\frac{dS}{dT}\right)_p dT, \qquad dv = \left(\frac{dv}{dp}\right)_T dp + \left(\frac{dv}{dT}\right)_p dT$$

with
$$dU = \left[T\left(\frac{dS}{dp}\right)_T - p\left(\frac{dv}{dp}\right)_T\right]dp + \left[T\left(\frac{dS}{dT}\right)_p - p\left(\frac{dv}{dT}\right)_p\right]dT$$

should be made. The expression of the condition is then

$$\left\{\frac{d}{dT}\left[T\left(\frac{dS}{dp}\right)_T - p\left(\frac{dv}{dp}\right)_T\right]\right\}_p = \left\{\frac{d}{dp}\left[T\left(\frac{dS}{dT}\right)_p - p\left(\frac{dv}{dT}\right)_p\right]\right\}_T$$

or
$$\left(\frac{dS}{dp}\right)_T + T\frac{\partial^2 S}{\partial T\partial p} - p\frac{\partial^2 v}{\partial T\partial p} = T\frac{\partial^2 S}{\partial p\partial T} - \left(\frac{dv}{dT}\right)_p - p\frac{\partial^2 v}{\partial p\partial T},$$

where the differentiation on the left is made with p constant and that on the right with T constant and where the subscripts have been dropped from the second derivatives and the usual notation adopted. Everything cancels except two terms which give

$$\left(\frac{dS}{dp}\right)_T = -\left(\frac{dv}{dT}\right)_p \quad \text{or} \quad \frac{1}{T}\left(\frac{TdS}{dp}\right)_T = -\left(\frac{dv}{dT}\right)_p. \tag{28}$$

The importance of the test for an exact differential lies not only in the relations obtained between the derivatives as above, but also in the fact that in applied mathematics a great many expressions are written as differentials which are not the total differentials of any functions and which must be distinguished from exact differentials. For instance if dH denote the infinitesimal portion of heat added to the gas or vapor above considered, the fundamental equation is expressed as $dH = dU + pdv$. That is to say, the amount of heat added is equal to the increase in the energy plus the work done by the gas in expanding. Now dH is not the differential of any function $H(U, v)$; it is $dS = dH/T$ which is the differential, and this is one reason for introducing the entropy S. Again if the forces X, Y act on a particle, the *work* done during the displacement through the arc $ds = \sqrt{dx^2 + dy^2}$ is written $dW = Xdx + Ydy$. It may happen that this is the total differential of some function; indeed, if

$$dW = -dV(x, y), \quad Xdx + Ydy = -dV, \quad X = -\frac{\partial V}{\partial x}, \quad Y = -\frac{\partial V}{\partial y},$$

where the negative sign is introduced in accordance with custom, the function V is called the *potential energy* of the particle. In general, however, there is no potential energy function V, and dW is not an exact differential; this is always true when part of the work is due to forces of friction. A notation which should distinguish between exact differentials and those which are not exact is much more needed than a notation to distinguish between partial and ordinary derivatives; but there appears to be none.

Many of the physical magnitudes of thermodynamics are expressed as derivatives and such relations as (26) establish relations between the magnitudes. Some definitions:

specific heat at constant volume is $C_v = \left(\frac{dH}{dT}\right)_v = T\left(\frac{dS}{dT}\right)_v$,

specific heat at constant pressure is $C_p = \left(\frac{dH}{dT}\right)_p = T\left(\frac{dS}{dT}\right)_p$,

latent heat of expansion is $L_v = \left(\frac{dH}{dv}\right)_T = T\left(\frac{dS}{dv}\right)_T$,

coefficient of cubic expansion is $\alpha_p = \frac{1}{v}\left(\frac{dv}{dT}\right)_p$,

modulus of elasticity (isothermal) is $E_T = -v\left(\frac{dp}{dv}\right)_T$,

modulus of elasticity (adiabatic) is $E_S = -v\left(\frac{dp}{dv}\right)_S$.

53. A polynomial is said to be homogeneous when each of its terms is of the same order when all the variables are considered. A definition of homogeneity which includes this case and is applicable to more general cases is: *A function $f(x, y, z, \cdots)$ of any number of variables is called homogeneous if the function is multiplied by some power of λ when all the variables are multiplied by λ*: and the power of λ which factors

out is called the order of homogeneity of the function. In symbols the condition for homogeneity of order n is

$$f(\lambda x, \lambda y, \lambda z, \cdots) = \lambda^n f(x, y, z, \cdots). \tag{29}$$

Thus
$$xe^{\frac{y}{x}} + \frac{y^2}{x}, \quad \frac{xy}{z^2} + \tan^{-1}\frac{x}{z}, \quad \frac{1}{\sqrt{x^2+y^2}} \tag{29'}$$

are homogeneous functions of order 1, 0, -1 respectively. To test a function for homogeneity it is merely necessary to replace all the variables by λ times the variables and see if λ factors out completely. The homogeneity may usually be seen without the test.

If the identity (29) be differentiated with respect to λ, with $x' = \lambda x$, etc.,

$$\left(x\frac{\partial}{\partial x'} + y\frac{\partial}{\partial y'} + z\frac{\partial}{\partial z'} + \cdots\right) f(\lambda x, \lambda y, \lambda z, \cdots) = n\lambda^{n-1} f(x, y, z, \cdots).$$

A second differentiation with respect to λ would give

$$\left(x^2\frac{\partial^2}{\partial x'^2} + xy\frac{\partial^2}{\partial x'\partial y'} + xz\frac{\partial^2}{\partial x'\partial z'} + \cdots\right)f + \left(yx\frac{\partial^2}{\partial y'\partial x'} + y^2\frac{\partial^2}{\partial y'^2} + yz\frac{\partial^2}{\partial y'\partial z'} + \cdots\right)f$$
$$+\left(zx\frac{\partial^2}{\partial z'\partial x'} + zy\frac{\partial^2}{\partial z'\partial y'} + z^2\frac{\partial^2}{\partial z'^2} + \cdots\right)f + \cdots = n(n-1)\lambda^{n-2} f(x, y, z, \cdots)$$

or
$$\left(x^2\frac{\partial^2}{\partial x'^2} + 2xy\frac{\partial^2}{\partial x'\partial y'} + y^2\frac{\partial^2}{\partial y'^2} + \cdots\right)f = n(n-1)\lambda^{n-2} f(x, y, z, \cdots).$$

Now if λ be set equal to 1 in these equations, then $x' = x$ and

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} + z\frac{\partial f}{\partial z} + \cdots = nf(x, y, z, \cdots), \tag{30}$$

$$x^2\frac{\partial^2 f}{\partial x^2} + 2xy\frac{\partial^2 f}{\partial x\partial y} + y^2\frac{\partial^2 f}{\partial y^2} + 2xz\frac{\partial^2 f}{\partial x\partial z} + \cdots = n(n-1)f(x, y, z, \cdots).$$

In words, these equations state that the sum of the partial derivatives each multiplied by the variable with respect to which the differentiation is performed is n times the function if the function is homogeneous of order n; and that the sum of the second derivatives each multiplied by the variables involved and by 1 or 2, according as the variable is repeated or not, is $n(n-1)$ times the function. The general formula obtained by differentiating any number of times with respect to λ may be expressed symbolically in the convenient form

$$(xD_x + yD_y + zD_z + \cdots)^k f = n(n-1)\cdots(n-k+1)f. \tag{31}$$

This is known as *Euler's Formula* on homogeneous functions.

It is worth while noting that in a certain sense every equation which represents a geometric or physical relation is homogeneous. For instance, in geometry the magnitudes that arise may be lengths, areas, volumes, or angles. These magnitudes are expressed as a number times a unit; thus, $\sqrt{2}$ ft., 3 sq. yd., π cu. ft.

In adding and subtracting, the terms must be like quantities; lengths added to lengths, areas to areas, etc. The *fundamental unit* is taken as length. The units of area, volume, and angle are *derived* therefrom. Thus the area of a rectangle or the volume of a rectangular parallelepiped is

$$A = a \text{ ft.} \times b \text{ ft.} = ab \text{ ft.}^2 = ab \text{ sq ft.}, \quad V = a \text{ ft.} \times b \text{ ft.} \times c \text{ ft.} = abc \text{ ft.}^3 = abc \text{ cu. ft.},$$

and the units sq. ft., cu. ft. are denoted as ft.^{2}, ft.^{3} just as if the simple unit ft. had been treated as a literal quantity and included in the multiplication. An area or volume is therefore considered as a compound quantity consisting of a number which gives its magnitude and a unit which gives its quality or dimensions. If L denote length and $[L]$ denote "of the dimensions of length," and if similar notations be introduced for area and volume, the equations $[A] = [L]^2$ and $[V] = [L]^3$ state that the dimensions of area are squares of length, and of volumes, cubes of lengths. If it be recalled that for purposes of analysis an angle is measured by the ratio of the arc subtended to the radius of the circle, the dimensions of angle are seen to be nil, as the definition involves the ratio of like magnitudes and must therefore be a *pure number*.

When geometric facts are represented analytically, either of two alternatives is open: 1°, the equations may be regarded as existing between mere numbers; or 2°, as between actual magnitudes. Sometimes one method is preferable, sometimes the other. Thus the equation $x^2 + y^2 = r^2$ of a circle may be interpreted as 1°, the sum of the squares of the coördinates (numbers) is constant; or 2°, the sum of the squares on the legs of a right triangle is equal to the square on the hypotenuse (Pythagorean Theorem). The second interpretation better sets forth the true inwardness of the equation. Consider in like manner the parabola $y^2 = 4px$. Generally y and x are regarded as mere numbers, but they may equally be looked upon as lengths and then the statement is that the square upon the ordinate equals the rectangle upon the abscissa and the constant length $4p$; this may be interpreted into an actual construction for the parabola, because a square equivalent to a rectangle may be constructed.

In the last interpretation the constant p was assigned the dimensions of length so as to render the equation homogeneous in dimensions, with each term of the dimensions of area or $[L]^2$. It will be recalled, however, that in the definition of the parabola, the quantity p actually has the dimensions of length, being half the distance from the fixed point to the fixed line (focus and directrix). This is merely another corroboration of the initial statement that the equations which actually arise in considering geometric problems are homogeneous in their dimensions, and must be so for the reason that in stating the first equation like magnitudes must be compared with like magnitudes.

The question of dimensions may be carried along through such processes as differentiation and integration. For let y have the dimensions $[y]$ and x the dimensions $[x]$. Then Δy, the difference of two y's, must still have the dimensions $[y]$ and Δx the dimensions $[x]$. The quotient $\Delta y/\Delta x$ then has the dimensions $[y]/[x]$. For example the relations for area and for volume of revolution,

$$\frac{dA}{dx} = y, \quad \frac{dV}{dx} = \pi y^2, \quad \text{give} \quad \left[\frac{dA}{dx}\right] = \frac{[A]}{[L]} = [L], \quad \left[\frac{dV}{dx}\right] = \frac{[V]}{[L]} = [L]^2,$$

and the dimensions of the left-hand side check with those of the right-hand side. As integration is the limit of a sum, the dimensions of an integral are the product

of the dimensions of the function to be integrated and of the differential dx. Thus if

$$y = \int_0^x \frac{dx}{a^2 + x^2} = \frac{1}{a}\tan^{-1}\frac{x}{a} + c$$

were an integral arising in actual practice, the very fact that a^2 and x^2 are added would show that they must have the same dimensions. If the dimensions of x be $[L]$, then

$$\left[\int_0^x \frac{dx}{a^2 + x^2}\right] = \left[\frac{1}{a^2 + x^2}\right][dx] = \frac{1}{[L]^2}[L] = \frac{1}{[L]} = [y],$$

and this checks with the dimensions on the right which are $[L]^{-1}$, since angle has no dimensions. As a rule, the theory of dimensions is neglected in pure mathematics; but it can nevertheless be made exceedingly useful and instructive.

In mechanics the *fundamental units* are length, mass, and time; and are denoted by $[L]$, $[M]$, $[T]$. The following table contains some derived units:

velocity	$\frac{[L]}{[T]}$,	acceleration	$\frac{[L]}{[T]^2}$,	force	$\frac{[M][L]}{[T]^2}$,
areal velocity	$\frac{[L]^2}{[T]}$,	density	$\frac{[M]}{[L]^3}$,	momentum	$\frac{[M][L]}{[T]}$,
angular velocity	$\frac{1}{[T]}$,	moment	$\frac{[M][L]^2}{[T]^2}$,	energy	$\frac{[M][L]^2}{[T]^2}$.

With the aid of a table like this it is easy to convert magnitudes in one set of units as ft., lb., sec., to another system, say cm., gm., sec. All that is necessary is to substitute for each individual unit its value in the new system. Thus

$$g = 32\tfrac{1}{6}\,\frac{\text{ft.}}{\text{sec.}^2}, \qquad 1\text{ ft.} = 30.48\text{ cm.}, \qquad g = 32\tfrac{1}{6} \times 30.48\,\frac{\text{cm.}}{\text{sec.}^2} = 980\tfrac{1}{2}\,\frac{\text{cm.}}{\text{sec.}^2}.$$

EXERCISES

1. Obtain the derivatives f''_{xx}, f''_{xy}, f''_{yx}, f''_{yy} and verify $f''_{xy} = f''_{yx}$.

$$(\alpha)\ \sin^{-1}\frac{y}{x}, \qquad (\beta)\ \log\frac{x^2 + y^2}{xy}, \qquad (\gamma)\ \phi\left(\frac{y}{x}\right) + \psi(xy).$$

2. Compute $\partial^2 v/\partial y^2$ in polar coördinates by the straightforward method.

3. Show that $a^2\dfrac{\partial^2 v}{\partial x^2} = \dfrac{\partial^2 v}{\partial t^2}$ if $v = f(x + at) + \phi(x - at)$.

4. Show that this equation is unchanged in form by the transformation:

$$\frac{\partial^2 f}{\partial x^2} + 2xy^2\frac{\partial f}{\partial x} + 2(y - y^3)\frac{\partial f}{\partial y} + x^2y^2 f = 0; \qquad u = xy, \quad v = 1/y.$$

5. In polar coördinates $z = r\cos\theta$, $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$ in space

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} = \frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial v}{\partial r}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 v}{\partial \phi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial v}{\partial \theta}\right)\right].$$

The work of transformation may be shortened by substituting successively

$$x = r_1\cos\phi, \quad y = r_1\sin\phi, \quad \text{and} \quad z = r\cos\phi, \quad r_1 = r\sin\phi.$$

6. Let x, y, z, t be four independent variables and $x = r\cos\phi$, $y = r\sin\phi$, $z = z$ the equations for transforming x, y, z to cylindrical coördinates. Let

$$X = -\frac{\partial^2 f}{\partial x \partial z}, \quad Y = -\frac{\partial^2 f}{\partial y \partial z}, \quad Z = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}, \quad F = \frac{\partial^2 f}{\partial y \partial t}, \quad G = -\frac{\partial^2 f}{\partial x \partial t};$$

show $\quad Z = \frac{1}{r}\frac{\partial Q}{\partial r}, \quad X\cos\phi + Y\sin\phi = -\frac{1}{r}\frac{\partial Q}{\partial z}, \quad F\sin\phi - G\cos\phi = \frac{1}{r}\frac{\partial Q}{\partial t},$

where $r^{-1}Q = \partial f/\partial r$. (Of importance for the Hertz oscillator.) Take $\partial f/\partial \phi = 0$.

7. Apply the test for an exact differential to each of the following, and write by inspection the functions corresponding to the exact differentials:

(α) $3xdx + y^2dy$, (β) $3xydx + x^3dy$, (γ) $x^2ydx + y^2dy$,

(δ) $\dfrac{xdx + ydy}{x^2 + y^2}$, ($\epsilon$) $\dfrac{xdx - ydy}{x^2 + y^2}$, ($\zeta$) $\dfrac{ydx - xdy}{x^2 + y^2}$,

(η) $(4x^3 + 3x^2y + y^3)dx + (x^3 + 2xy + 3y^2)dy$. ($\theta$) $x^2y^2(dx + dy)$.

8. Express the conditions that $P(x, y, z)dx + Q(x, y, z)dy + R(x, y, z)dz$ be an exact differential $dF(x, y, z)$. Apply these conditions to the differentials:

(α) $3x^2y^2zdx + 2x^3yzdy + x^3y^2dz$, ($\beta$) $(y + z)dx + (x + z)dy + (x + y)dz$.

9. Obtain $\left(\dfrac{dp}{dT}\right)_v = \left(\dfrac{dS}{dv}\right)_T$ and $\left(\dfrac{dv}{dS}\right)_p = \left(\dfrac{dT}{dp}\right)_S$ from (27) with proper variables.

10. If three functions (called thermodynamic potentials) be defined as

$$\psi = U - TS, \qquad \chi = U + pv, \qquad \zeta = U - TS + pv,$$

show $\qquad d\psi = -SdT - pdv, \quad d\chi = TdS + vdp, \quad d\zeta = -SdT + vdp,$

and express the conditions that $d\psi$, $d\chi$, $d\zeta$ be exact. Compare with Ex. 9.

11. State in words the definitions corresponding to the defining formulas, p. 107.

12. If the sum $(Mdx + Ndy) + (Pdx + Qdy)$ of two differentials is exact and one of the differentials is exact, the other is. Prove this.

13. Apply Euler's Formula (31), for the simple case $k = 1$, to the three functions (29′) and verify the formula. Apply it for $k = 2$ to the first function.

14. Verify the homogeneity of these functions and determine their order:

(α) $y^2/x + x(\log x - \log y)$, ($\beta$) $\dfrac{x^m y^n}{\sqrt{x^2 + y^2}}$, ($\gamma$) $\dfrac{xyz}{ax + by + cz}$,

(δ) $xye^{\frac{xz}{y^2}} + z^2$, ($\epsilon$) $\sqrt{x}\cot^{-1}\dfrac{y}{z}$, ($\zeta$) $\dfrac{\sqrt[5]{x} - \sqrt[5]{y}}{\sqrt[4]{x} + \sqrt[4]{y}}$.

15. State the dimensions of moment of inertia and convert a unit of moment of inertia in ft.-lb. into its equivalent in cm.-gm.

16. Discuss for dimensions Peirce's formulas Nos. 93, 124–125, 220, 300.

17. Continue Ex. 17, p. 101, to show $\dfrac{d}{dt}\dfrac{\partial \dot x}{\partial \dot q_i} = \dfrac{\partial \dot x}{\partial q_i}$ and $\dfrac{d}{dt}\dfrac{\partial T}{\partial \dot q_i} = m\ddot x\dfrac{\partial x}{\partial q_i} + \dfrac{\partial T}{\partial q_i}$.

18. If $p_i = \dfrac{\partial T}{\partial \dot q_i}$ in Ex. 17, p. 101, show without analysis that $2T = \dot q_1 p_1 + \dot q_2 p_2$. If T' denote $T' = T$, where T' is considered as a function of p_1, p_2 while T is considered as a function of $\dot q_1$, $\dot q_2$, prove from $T' = \dot q_1 p_1 + \dot q_2 p_2 - T$ that

$$\frac{\partial T'}{\partial p_i} = \dot q_i, \qquad \frac{\partial T'}{\partial q_i} = -\frac{\partial T}{\partial q_i}.$$

19. If (x_1, y_1) and (x_2, y_2) are the coördinates of two moving particles and

$$m_1\frac{d^2x_1}{dt^2} = X_1,\quad m_1\frac{d^2y_1}{dt^2} = Y_1,\quad m_2\frac{d^2x_2}{dt^2} = X_2,\quad m_2\frac{d^2y_2}{dt^2} = Y_2$$

are the equations of motion, and if x_1, y_1, x_2, y_2 are expressible as

$$x_1 = f_1(q_1, q_2, q_3),\quad y_1 = g_1(q_1, q_2, q_3),\quad x_2 = f_2(q_1, q_2, q_3),\quad y_2 = g_2(q_1, q_2, q_3)$$

in terms of three independent variables q_1, q_2, q_3, show that

$$Q_1 = X_1\frac{\partial x_1}{\partial q_1} + Y_1\frac{\partial y_1}{\partial q_1} + X_2\frac{\partial x_2}{\partial q_1} + Y_2\frac{\partial y_2}{\partial q_1} = \frac{d}{dt}\frac{\partial T}{\partial \dot q_1} - \frac{\partial T}{\partial q_1},$$

where $T = \frac{1}{2}(m_1v_1^2 + m_2v_2^2) = T(q_1, q_2, q_3, \dot q_1, \dot q_2, \dot q_3)$ and is homogeneous of the second degree in $\dot q_1, \dot q_2, \dot q_3$. The work may be carried on as a generalization of Ex. 17, p. 101, and Ex. 17 above. It may be further extended to any number of particles whose positions in space depend on a number of variables q.

20. In Ex. 19 if $p_i = \dfrac{\partial T}{\partial \dot q_i}$, generalize Ex. 18 to obtain

$$\dot q_i = \frac{\partial T'}{\partial p_i},\quad \frac{\partial T'}{\partial q_i} = -\frac{\partial T}{\partial q_i},\quad Q_1 = \frac{dp_1}{dt} + \frac{\partial T'}{\partial q_1}.$$

The equations $Q_i = \dfrac{d}{dt}\dfrac{\partial T}{\partial \dot q_i} - \dfrac{\partial T}{\partial q_i}$ and $Q_i = \dfrac{dp_i}{dt} + \dfrac{\partial T'}{\partial q_i}$ are respectively the Lagrangian and Hamiltonian equations of motion.

21. If $rr' = k^2$ and $\phi' = \phi$ and $v'(r', \phi') = v(r, \phi)$, show

$$\frac{\partial^2 v'}{\partial r'^2} + \frac{1}{r'}\frac{\partial v'}{\partial r'} + \frac{1}{r'^2}\frac{\partial^2 v'}{\partial \phi'^2} = \frac{r^2}{r'^2}\left(\frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} + \frac{1}{r^2}\frac{\partial^2 v}{\partial \phi^2}\right).$$

22. If $rr' = k^2$, $\phi' = \phi$, $\theta' = \theta$, and $v'(r', \phi', \theta') = \dfrac{k}{r'}v(r, \phi, \theta)$, show that the expression of Ex. 5 in the primed letters is kr^2/r'^3 of its value for the unprimed letters. (Useful in § 198.)

23. If $z = x\phi\left(\dfrac{y}{x}\right) + \psi\left(\dfrac{y}{x}\right)$, show $x^2\dfrac{\partial^2 z}{\partial x^2} + 2xy\dfrac{\partial^2 z}{\partial x\partial y} + y^2\dfrac{\partial^2 z}{\partial y^2} = 0.$

24. Make the indicated changes of variable:

(α) $\dfrac{\partial^2 V}{\partial x^2} + \dfrac{\partial^2 V}{\partial y^2} = e^{-2u}\left(\dfrac{\partial^2 V}{\partial u^2} + \dfrac{\partial^2 V}{\partial v^2}\right)$ if $x = e^u\cos v$, $y = e^u\sin v$,

(β) $\dfrac{\partial^2 V}{\partial u^2} + \dfrac{\partial^2 V}{\partial v^2} = \left(\dfrac{\partial^2 V}{\partial x^2} + \dfrac{\partial^2 V}{\partial y^2}\right)\left[\left(\dfrac{\partial f}{\partial u}\right)^2 + \left(\dfrac{\partial f}{\partial v}\right)^2\right]$, where

$$x = f(u, v),\quad y = \phi(u, v),\quad \frac{\partial f}{\partial u} = \frac{\partial \phi}{\partial v},\quad \frac{\partial f}{\partial v} = -\frac{\partial \phi}{\partial u}.$$

25. For an orthogonal transformation (Ex. 10 (ζ), p. 100)

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} = \frac{\partial^2 v}{\partial x'^2} + \frac{\partial^2 v}{\partial y'^2} + \frac{\partial^2 v}{\partial z'^2}.$$

54. Taylor's Formula and applications. The development of $f(x, y)$ is found, as was the Theorem of the Mean, from the relation (p. 95)

$$\Delta f = \Phi(1) - \Phi(0) \quad \text{if} \quad \Phi(t) = f(a + th, b + tk).$$

If $\Phi(t)$ be expanded by Maclaurin's Formula to n terms,

$$\Phi(t) - \Phi(0) = t\Phi'(0) + \frac{t^2}{2!}\Phi''(0) + \cdots + \frac{t^{n-1}}{(n-1)!}\Phi^{(n-1)}(0) + \frac{t^n}{n!}\Phi^{(n)}(\theta t).$$

The expressions for $\Phi'(t)$ and $\Phi'(0)$ may be found as follows by (10):

$$\Phi'(t) = hf'_x + kf'_y, \quad \Phi'(0) = [hf'_x + kf'_y]_{\substack{x=a\\y=b}},$$

then
$$\Phi''(t) = h(hf''_{xx} + kf''_{yx}) + k(hf''_{xy} + kf''_{yy})$$
$$= h^2f''_{xx} + 2hkf''_{xy} + k^2f''_{yy} = (hD_x + kD_y)^2 f,$$
$$\Phi^{(i)}(t) = (hD_x + hD_y)^i f, \quad \Phi^{(i)}(0) = [(hD_x + kD_y)^i f]_{\substack{x=a\\y=b}}.$$

And $f(a + h, b + k) - f(a, b) = \Delta f = \Phi(1) - \Phi(0) = (hD_x + kD_y)f(a, b)$

$$+ \frac{1}{2!}(hD_x + kD_y)^2 f(a, b) + \cdots + \frac{1}{(n-1)!}(hD_x + kD_y)^{n-1} f(a, b)$$
$$+ \frac{1}{n!}(hD_x + kD_y)^n f(a + \theta h, b + \theta k). \tag{32}$$

In this expansion, the increments h and k may be replaced, if desired, by $x - a$ and $y - b$ and then $f(x, y)$ will be expressed in terms of its value and the values of its derivatives at (a, b) in a manner entirely analogous to the case of a single variable. In particular if the point (a, b) about which the development takes place be (0, 0) the development becomes Maclaurin's Formula for $f(x, y)$.

$$f(x, y) = f(0, 0) + (xD_x + yD_y)f(0, 0) + \frac{1}{2!}(xD_x + yD_y)^2 f(0, 0) + \cdots$$
$$+ \frac{1}{(n-1)!}(xD_x + yD_y)^{n-1} f(0, 0) + \frac{1}{n!}(xD_x + yD_y)^n f(\theta x, \theta y). \tag{32'}$$

Whether in Maclaurin's or Taylor's Formula, the successive terms are homogeneous polynomials of the 1st, 2d, $\cdots$, $(n - 1)$st order in x, y or in $x - a$, $y - b$. The formulas are unique as in § 32.

Suppose $\sqrt{1 - x^2 - y^2}$ is to be developed about (0, 0). The successive derivatives are

$$f'_x = \frac{-x}{\sqrt{1 - x^2 - y^2}}, \quad f'_y = \frac{-y}{\sqrt{1 - x^2 - y^2}}, \quad f'_x(0, 0) = 0, \quad f'_y(0, 0) = 0,$$

$$f''_{xx} = \frac{-1 + y^2}{(1 - x^2 - y^2)^{\frac{3}{2}}}, \quad f''_{xy} = \frac{xy}{(1 - x^2 - y^2)^{\frac{3}{2}}}, \quad f''_{yy} = \frac{-1 + x^2}{(1 - x^2 - y^2)^{\frac{3}{2}}},$$

$$f'''_{x^3} = \frac{3(1 - y^2)x}{(1 - x^2 - y^2)^{\frac{5}{2}}}, \quad f'''_{x^2y} = \frac{y^3 - 2xy^2 - y}{(1 - x^2 - y^2)^{\frac{5}{2}}}, \quad \cdots,$$

and $\sqrt{1 - x^2 - y^2} = 1 + (0x + 0y) + \frac{1}{2}(-x^2 + 0xy - y^2) + \frac{1}{6}(0x^3 + \cdots) + \cdots,$

or $\sqrt{1 - x^2 - y^2} = 1 - \frac{1}{2}(x^2 + y^2) + \text{terms of fourth order} + \cdots.$

In this case the expansion may be found by treating $x^2 + y^2$ as a single term and expanding by the binomial theorem. The result would be

$$[1-(x^2+y^2)]^{\frac{1}{2}} = 1 - \tfrac{1}{2}(x^2+y^2) - \tfrac{1}{8}(x^4+2x^2y^2+y^4) - \tfrac{1}{16}(x^2+y^2)^3 - \cdots.$$

That the development thus obtained is identical with the Maclaurin development that might be had by the method above, follows from the uniqueness of the development. Some such short cut is usually available.

55. The condition that a function $z = f(x, y)$ have a minimum or maximum at (a, b) is that $\Delta f > 0$ or $\Delta f < 0$ for all values of $h = \Delta x$ and $k = \Delta y$ which are sufficiently small. From either geometrical or analytic considerations it is seen that if the surface $z = f(x, y)$ has a minimum or maximum at (a, b), the curves in which the planes $y = b$ and $x = a$ cut the surface have minima or maxima at $x = a$ and $y = b$ respectively. Hence the partial derivatives f'_x and f'_y must both vanish at (a, b), provided, of course, that exceptions like those mentioned on page 7 be made. The two simultaneous equations

$$f'_x = 0, \quad f'_y = 0, \tag{33}$$

corresponding to $f'(x) = 0$ in the case of a function of a single variable, may then be solved to find the positions (x, y) of the minima and maxima. Frequently the geometric or physical interpretation of $z = f(x, y)$ or some special device will then determine whether there is a maximum or a minimum or neither at each of these points.

For example let it be required to find the maximum rectangular parallelepiped which has three faces in the coördinate planes and one vertex in the plane $x/a + y/b + z/c = 1$. The volume is

$$V = xyz = cxy\left(1 - \frac{x}{a} - \frac{y}{b}\right).$$

$$\frac{\partial V}{\partial x} = -2\frac{c}{a}xy - \frac{c}{b}y^2 + cy = 0 \qquad \frac{\partial V}{\partial y} = -2\frac{c}{b}xy - \frac{c}{a}x^2 + cx = 0.$$

The solution of these equations is $x = \frac{1}{3}a$, $y = \frac{1}{3}b$. The corresponding z is $\frac{1}{3}c$ and the volume V is therefore $abc/9$ or $\frac{2}{3}$ of the volume cut off from the first octant by the plane. It is evident that this solution is a maximum. There are other solutions of $V'_x = V'_y = 0$ which have been discarded because they give $V = 0$.

The conditions $f'_x = f'_y = 0$ may be established analytically. For

$$\Delta f = (f'_x + \zeta_1)\Delta x + (f'_y + \zeta_2)\Delta y.$$

Now as ζ_1, ζ_2 are infinitesimals, the signs of the parentheses are determined by the signs of f'_x, f'_y unless these derivatives vanish; and hence unless $f'_x = 0$, the sign of Δf for Δx sufficiently small and positive and $\Delta y = 0$ would be opposite to the sign of Δf for Δx sufficiently small and negative and $\Delta y = 0$. Therefore *for a minimum or maximum $f'_x = 0$; and in like manner $f'_y = 0$*. Considerations like these will serve to establish a criterion for distinguishing between maxima and minima

analogous to the criterion furnished by $f''(x)$ in the case of one variable. For if $f'_x = f'_y = 0$, then

$$\Delta f = \tfrac{1}{2}\left(h^2 f''_{xx} + 2\,hk f''_{xy} + k^2 f''_{yy}\right)_{x = a + \theta h,\ y = b + \theta k},$$

by Taylor's Formula to two terms. Now if the second derivatives are continuous functions of (x, y) in the neighborhood of (a, b), each derivative at $(a + \theta h, b + \theta k)$ may be written as its value at (a, b) plus an infinitesimal. Hence

$$\Delta f = \tfrac{1}{2}\left(h^2 f''_{xx} + 2\,hk f''_{xy} + k^2 f''_{yy}\right)_{(a,\,b)} + \tfrac{1}{2}\left(h^2 \zeta_1 + 2\,hk\zeta_2 + k^2\zeta_3\right).$$

Now the sign of Δf for sufficiently small values of h, k must be the same as the sign of the first parenthesis provided that parenthesis does not vanish. Hence if the quantity

$$\left(h^2 f''_{xx} + 2\,hk f''_{xy} + k^2 f''_{yy}\right)_{(a,\,b)} \begin{matrix} > 0 \text{ for every } (h, k), \text{ a minimum} \\ < 0 \text{ for every } (h, k), \text{ a maximum.} \end{matrix}$$

As the derivatives are taken at the point (a, b), they have certain constant values, say A, B, C. The question of distinguishing between minima and maxima therefore reduces to the discussion of the possible signs of a *quadratic form* $Ah^2 + 2\,Bhk + Ck^2$ for different values of h and k. The examples

$$h^2 + k^2, \quad -h^2 - k^2, \quad h^2 - k^2, \quad \pm(h - k)^2$$

show that a quadratic form may be: either 1°, positive for every (h, k) except $(0, 0)$; or 2°, negative for every (h, k) except $(0, 0)$; or 3°, positive for some values (h, k) and negative for others and zero for others; or finally 4°, zero for values other than $(0, 0)$, but either never negative or never positive. Moreover, the four possibilities here mentioned are the only cases conceivable except 5°, that $A = B = C = 0$ and the form always is 0. In the first case the form is called a *definite positive* form, in the second a *definite negative* form, in the third an *indefinite* form, and in the fourth and fifth a *singular* form. The first case assures a minimum, the second a maximum, the third neither a minimum nor a maximum (sometimes called a minimax); but the case of a singular form leaves the question entirely undecided just as the condition $f''(x) = 0$ did.

The conditions which distinguish between the different possibilities may be expressed in terms of the coefficients A, B, C.

$$\begin{array}{llll} 1° \text{ pos. def.,} & B^2 < AC, \quad A, C > 0; & 3° \text{ indef.,} & B^2 > AC; \\ 2° \text{ neg. def.,} & B^2 < AC, \quad A, C < 0; & 4° \text{ sing.,} & B^2 = AC. \end{array}$$

The conditions for distinguishing between maxima and minima are:

$$\left.\begin{matrix} f'_x = 0 \\ f'_y = 0 \end{matrix}\right\} \quad f''^2_{xy} < f''_{xx} f''_{yy}, \quad \left\{\begin{matrix} f''_{xx} + f''_{yy} > 0 \text{ minimum}; \\ f''_{xx} + f''_{yy} < 0 \text{ maximum}; \end{matrix}\right. \tag{34}$$

$$f''^2_{xy} > f''_{xx} f''_{yy}, \text{ minimax}; \quad f''^2_{xy} = f''_{xx} f''_{yy} \text{ (?)}.$$

It may be noted that in applying these conditions to the case of a definite form it is sufficient to show that either f''_{xx} or f''_{yy} is positive or negative because they necessarily have the same sign.

EXERCISES

1. Write at length, without symbolic shortening, the expansion of $f(x, y)$ by Taylor's Formula to and including the terms of the third order in $x-a$, $y-b$. Write the formula also with the terms of the third order as the remainder.

2. Write by analogy the proper form of Taylor's Formula for $f(x, y, z)$ and prove it. Indicate the result for any number of variables.

3. Obtain the quadratic and lower terms in the development

$$(\alpha) \text{ of } xy^2 + \sin xy \text{ at } (1, \tfrac{1}{2}\pi) \quad \text{and} \quad (\beta) \text{ of } \tan^{-1}(y/x) \text{ at } (1, 1).$$

4. A rectangular parallelepiped with one vertex at the origin and three faces in the coördinate planes has the opposite vertex upon the ellipsoid

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1.$$

Find the maximum volume.

5. Find the point within a triangle such that the sum of the squares of its distances to the vertices shall be a minimum. Note that the point is the intersection of the medians. Is it obvious that a minimum and not a maximum is present?

6. A floating anchorage is to be made with a cylindrical body and equal conical ends. Find the dimensions that make the surface least for a given volume.

7. A cylindrical tent has a conical roof. Find the best dimensions.

8. Apply the test by second derivatives to the problem in the text and to any of Exs. 4–7. Discuss for maxima or minima the following functions:

$$(\alpha)\ x^2y + xy^2 - x, \qquad (\beta)\ x^3 + y^3 - x^2y^2 - \tfrac{1}{2}(x^2 + y^2),$$
$$(\gamma)\ x^2 + y^2 + x + y, \qquad (\delta)\ \tfrac{1}{3}y^3 - xy^2 + x^2y - x,$$
$$(\epsilon)\ x^3 + y^3 - 9xy + 27, \qquad (\zeta)\ x^4 + y^4 - 2x^2 + 4xy - 2y^2.$$

9. State the conditions on the first derivatives for a maximum or minimum of function of three or any number of variables. Prove in the case of three variables.

10. A wall tent with rectangular body and gable roof is to be so constructed as to use the least amount of tenting for a given volume. Find the dimensions.

11. Given any number of masses $m_1, m_2, \cdots, m_n$ situated at (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n). Show that the point about which their moment of inertia is least is their center of gravity. If the points were $(x_1, y_1, z_1), \cdots$ in space, what point would make Σmr^2 a minimum?

12. A test for maximum or minimum analogous to that of Ex. 27, p. 10, may be given for a function $f(x, y)$ of two variables, namely: If a function is positive all over a region and vanishes upon the contour of the region, it must have a maximum within the region at the point for which $f'_x = f'_y = 0$. If a function is finite all over a region and becomes infinite over the contour of the region, it must have a minimum within the region at the point for which $f'_x = f'_y = 0$. These tests are subject to the proviso that $f'_x = f'_y = 0$ has only a single solution. Comment on the test and apply it to exercises above.

13. If a, b, c, r are the sides of a given triangle and the radius of the inscribed circle, the pyramid of altitude h constructed on the triangle as base will have its maximum surface when the surface is $\frac{1}{2}(a+b+c)\sqrt{r^2+h^2}$.

CHAPTER V

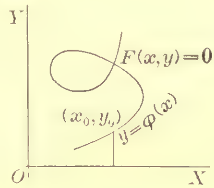
PARTIAL DIFFERENTIATION; IMPLICIT FUNCTIONS

56. The simplest case; $F(x, y) = 0$. The total differential

$$dF = F'_x dx + F'_y dy = d0 = 0$$

indicates
$$\frac{dy}{dx} = -\frac{F'_x}{F'_y}, \qquad \frac{dx}{dy} = -\frac{F'_y}{F'_x} \tag{1}$$

as the derivative of y by x, or of x by y, where y is defined as a function of x, or x as a function of y, by the relation $F(x, y) = 0$; and this method of obtaining a derivative of an *implicit function* without solving explicitly for the function has probably been familiar long before the notion of a partial derivative was obtained. The relation $F(x, y) = 0$ is pictured as a curve, and the function $y = \phi(x)$, which would be obtained by solution, is considered as multiple valued or as restricted to some definite portion or branch of the curve $F(x, y) = 0$. If the results (1) are to be applied to find the derivative at some point (x_0, y_0) of the curve $F(x, y) = 0$, it is necessary that at that point the denominator F'_y or F'_x should not vanish.



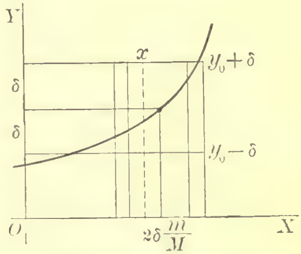
These pictorial and somewhat vague notions may be stated precisely as a *theorem* susceptible of proof, namely: Let x_0 be any real value of x such that 1°, the equation $F(x_0, y) = 0$ has a real solution y_0; and 2°, the function $F(x, y)$ regarded as a function of two independent variables (x, y) is continuous and has continuous first partial derivatives F'_x, F'_y in the neighborhood of (x_0, y_0); and 3°, the derivative $F'_y(x_0, y_0) \neq 0$ does not vanish for (x_0, y_0); then $F(x, y) = 0$ may be solved (theoretically) as $y = \phi(x)$ in the vicinity of $x = x_0$ and in such a manner that $y_0 = \phi(x_0)$, that $\phi(x)$ is continuous in x, and that $\phi(x)$ has a derivative $\phi'(x) = -F'_x/F'_y$; and the solution is unique. This is the fundamental theorem on implicit functions for the simple case, and the proof follows.

By the conditions on F'_x, F'_y, the Theorem of the Mean is applicable. Hence

$$F(x, y) - F(x_0, y_0) = F(x, y) = (hF'_x + kF'_y)_{x_0 + \theta h,\, y_0 + \theta k}. \tag{2}$$

Furthermore, in any square $|h| < \delta$, $|k| < \delta$ surrounding (x_0, y_0) and sufficiently small, the continuity of F'_x insures $|F'_x| < M$ and the continuity of F'_y taken with

the fact that $F'_y(x_0, y_0) \neq 0$ insures $|F'_y| > m$. Consider the range of x as further restricted to values such that $|x - x_0| < m\delta/M$ if $m < M$. Now consider the value of $F(x, y)$ for any x in the permissible interval and for $y = y_0 + \delta$ or $y = y_0 - \delta$. As $|kF'_y| > m\delta$ but $|(x - x_0)F'_x| < m\delta$, it follows from (2) that $F(x, y_0 + \delta)$ has the sign of $\delta F'_y$ and $F(x, y_0 - \delta)$ has the sign of $-\delta F'_y$; and as the sign of F'_y does not change, $F(x, y_0 + \delta)$ and $F(x, y_0 - \delta)$ have opposite signs. Hence by Ex. 10, p. 45, there is one and only one value of y between $y_0 - \delta$ and $y_0 + \delta$ such that $F(x, y) = 0$. Thus for each x in the interval there is one and only one y such that $F(x, y) = 0$. The equation $F(x, y) = 0$ has a unique solution near (x_0, y_0). Let $y = \phi(x)$ denote the solution. The solution is continuous at $x = x_0$ because $|y - y_0| < \delta$. If (x, y) are restricted to values $y = \phi(x)$ such that $F(x, y) = 0$, equation (2) gives at once



$$\frac{k}{h} = \frac{y - y_0}{x - x_0} = \frac{\Delta y}{\Delta x} = -\frac{F'_x(x + \theta h, y + \theta k)}{F'_y(x + \theta h, y + \theta k)}, \qquad \frac{dy}{dx} = -\frac{F'_x(x_0, y_0)}{F'_y(x_0, y_0)}.$$

As F'_x, F'_y are continuous and $F'_y \neq 0$, the fraction k/h approaches a limit and the derivative $\phi'(x_0)$ exists and is given by (1). The same reasoning would apply to any point x in the interval. The theorem is completely proved. It may be added that the expression for $\phi'(x)$ is such as to show that $\phi'(x)$ itself is continuous.

The values of higher derivatives of implicit functions are obtainable by successive total differentiation as

$$F'_x + F'_y y' = 0,$$

$$F''_{xx} + 2F''_{xy}y' + F''_{yy}y'^2 + F'_y y'' = 0, \tag{3}$$

etc. It is noteworthy that these successive equations may be solved for the derivative of highest order by dividing by F'_y which has been assumed not to vanish. The question of whether the function $y = \phi(x)$ defined implicitly by $F(x, y) = 0$ has derivatives of order higher than the first may be seen by these equations to depend on whether $F(x, y)$ has higher partial derivatives which are continuous in (x, y).

57. To find the *maxima and minima* of $y = \phi(x)$, that is, to find the points where the tangent to $F(x, y) = 0$ is parallel to the x-axis, observe that at such points $y' = 0$. Equations (3) give

$$F'_x = 0, \qquad F''_{xx} + F'_y y'' = 0. \tag{4}$$

Hence always under the assumption that $F'_y \neq 0$, *there are maxima at the intersections of $F = 0$ and $F'_x = 0$ if F''_{xx} and F'_y have the same sign, and minima at the intersections for which F''_{xx} and F'_y have opposite signs*; the case $F''_{xx} = 0$ still remains undecided.

For example if $F(x, y) = x^3 + y^3 - 3axy = 0$, the derivatives are

$$3(x^2 - ay) + 3(y^2 - ax)y' = 0, \qquad \frac{dy}{dx} = -\frac{x^2 - ay}{y^2 - ax},$$

$$6x - 6ay' + 6yy'^2 + 3(y^2 - ax)y'' = 0, \qquad \frac{d^2y}{dx^2} = -\frac{2a^3xy}{(y^2 - ax)^3}.$$

To find the maxima or minima of y as a function of x, solve

$$F'_x = 0 = x^2 - ay, \qquad F = 0 = x^3 + y^3 - 3axy, \qquad F'_y \neq 0.$$

The real solutions of $F'_x = 0$ and $F = 0$ are $(0, 0)$ and $(\sqrt[3]{2}a, \sqrt[3]{4}a)$ of which the first must be discarded because $F'_y(0, 0) = 0$. At $(\sqrt[3]{2}a, \sqrt[3]{4}a)$ the derivatives F'_y and F''_{xx} are positive; and the point is a maximum. The curve $F = 0$ is the folium of Descartes.

The rôle of the variables x and y may be interchanged if $F'_x \neq 0$ and the equation $F(x, y) = 0$ may be solved for $x = \psi(y)$, the functions ϕ and ψ being inverse. In this way the vertical tangents to the curve $F = 0$ may be discussed. For the points of $F = 0$ at which both $F'_x = 0$ and $F'_y = 0$, the equation cannot be solved in the sense here defined. Such points are called ***singular points*** of the curve. The questions of the singular points of $F = 0$ and of maxima, minima, or minimax (§ 55) of the surface $z = F(x, y)$ are related. For if $F'_x = F'_y = 0$, the surface has a tangent plane parallel to $z = 0$, and if the condition $z = F = 0$ is also satisfied, the surface is tangent to the xy-plane. Now if $z = F(x, y)$ has a maximum or minimum at its point of tangency with $z = 0$, the surface lies entirely on one side of the plane and the point of tangency is an isolated point of $F(x, y) = 0$; whereas if the surface has a minimax it cuts through the plane $z = 0$ and the point of tangency is not an isolated point of $F(x, y) = 0$. The shape of the curve $F = 0$ in the neighborhood of a singular point is discussed by developing $F(x, y)$ about that point by Taylor's Formula.

For example, consider the curve $F(x, y) = x^3 + y^3 - x^2y^2 - \frac{1}{2}(x^2 + y^2) = 0$ and the surface $z = F(x, y)$. The common real solutions of

$$F'_x = 3x^2 - 2xy^2 - x = 0, \qquad F'_y = 3y^2 - 2x^2y - y = 0, \qquad F(x, y) = 0$$

are the singular points. The real solutions of $F'_x = 0$, $F'_y = 0$ are $(0, 0)$, $(1, 1)$, $(\frac{1}{2}, \frac{1}{2})$ and of these the first two satisfy $F(x, y) = 0$ but the last does not. The singular points of the curve are therefore $(0, 0)$ and $(1, 1)$. The test (34) of § 55 shows that $(0, 0)$ is a maximum for $z = F(x, y)$ and hence an isolated point of $F(x, y) = 0$. The test also shows that $(1, 1)$ is a minimax. To discuss the curve $F(x, y) = 0$ near $(1, 1)$ apply Taylor's Formula.

$$0 = F(x, y) = \tfrac{1}{2}(3h^2 - 8hk + 3k^2) + \tfrac{1}{6}(6h^3 - 12h^2k - 12hk^2 + 6k^3) + \text{remainder}$$
$$= \tfrac{1}{2}(3\cos^2\phi - 8\sin\phi\cos\phi + 3\sin^2\phi)$$
$$+ r(\cos^3\phi - 2\cos^2\phi\sin\phi - 2\cos\phi\sin^2\phi + \sin^3\phi) + \cdots.$$

if polar coördinates $h = r\cos\phi$, $k = r\sin\phi$ be introduced at (1, 1) and r^2 be canceled. Now for very small values of r, the equation can be satisfied only when the first parenthesis is very small. Hence the solutions of

$$3 - 4\sin 2\phi = 0, \qquad \sin 2\phi = \tfrac{3}{4}, \qquad \text{or} \qquad \phi = 24^\circ 17\tfrac{1}{2}', \ 65^\circ 42\tfrac{1}{2}',$$

and $\phi + \pi$, are the directions of the tangents to $F(x, y) = 0$. The equation $F = 0$ is

$$0 = (1\tfrac{1}{2} - 2\sin 2\phi) + r(\cos\phi + \sin\phi)(1 - 1\tfrac{1}{2}\sin 2\phi)$$

if only the first two terms are kept, and this will serve to sketch $F(x, y) = 0$ for very small values of r, that is, for ϕ very near to the tangent directions.

58. It is important to obtain conditions for the maximum or minimum of a function $z = f(x, y)$ where the variables x, y are connected by a relation $F(x, y) = 0$ so that z really becomes a function of x alone or y alone. For it is not always possible, and frequently it is inconvenient, to solve $F(x, y) = 0$ for either variable and thus eliminate that variable from $z = f(x, y)$ by substitution. When the variables x, y in $z = f(x, y)$ are thus connected, the minimum or maximum is called a *constrained minimum* or *maximum*; when there is no equation $F(x, y) = 0$ between them the minimum or maximum is called *free* if any designation is needed.* The conditions are obtained by differentiating $z = f(x, y)$ and $F(x, y) = 0$ totally with respect to x. Thus

$$\frac{dz}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0, \qquad \frac{d0}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0,$$

and
$$\frac{\partial f}{\partial x}\frac{\partial F}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial F}{\partial x} = 0, \qquad \frac{d^2z}{dx^2} \gtrless 0, \qquad F = 0, \tag{5}$$

where the first equation arises from the two above by eliminating dy/dx and the second is added to insure a minimum or maximum, are the conditions desired. Note that all singular points of $F(x, y) = 0$ satisfy the first condition identically, but that the process by means of which it was obtained excludes such points, and that the rule cannot be expected to apply to them.

Another method of treating the problem of constrained maxima and minima is to introduce *a multiplier* and form the function

$$z = \Phi(x, y) = f(x, y) + \lambda F(x, y), \qquad \lambda \text{ a multiplier.} \tag{6}$$

Now if this function z is to have a free maximum or minimum, then

$$\Phi'_x = f'_x + \lambda F'_x = 0, \qquad \Phi'_y = f'_y + \lambda F'_y = 0. \tag{7}$$

These two equations taken with $F = 0$ constitute a set of three from which the three values x, y, λ may be obtained by solution. Note that

* The adjective "relative" is sometimes used for constrained, and "absolute" for free; but the term "absolute" is best kept for the greatest of the maxima or least of the minima, and the term "relative" for the other maxima and minima.

λ cannot be obtained from (7) if both F'_x and F'_y vanish; and hence this method also rejects the singular points. That this method really determines the constrained maxima and minima of $f(x, y)$ subject to the constraint $F(x, y) = 0$ is seen from the fact that if λ be eliminated from (7) the condition $f'_x F'_y - f'_y F'_x = 0$ of (5) is obtained. The new method is therefore identical with the former, and its introduction is more a matter of convenience than necessity. It is possible to show directly that the new method gives the constrained maxima and minima. For the conditions (7) are those of a free extreme for the function $\Phi(x, y)$ which depends on two independent variables (x, y). Now if the equations (7) be solved for (x, y), it appears that the position of the maximum or minimum will be expressed in terms of λ as a parameter and that consequently the point $(x(\lambda), y(\lambda))$ cannot in general lie on the curve $F(x, y) = 0$; but if λ be so determined that the point shall lie on this curve, the function $\Phi(x, y)$ has a free extreme at a point for which $F = 0$ and hence in particular must have a constrained extreme for the particular values for which $F(x, y) = 0$. In speaking of (7) as the conditions for an extreme, the conditions which should be imposed on the second derivative have been disregarded.

For example, suppose the maximum radius vector from the origin to the folium of Descartes were desired. The problem is to render $f(x, y) = x^2 + y^2$ maximum subject to the condition $F(x, y) = x^3 + y^3 - 3axy = 0$. Hence

$$2x + 3\lambda(x^2 - ay) = 0, \qquad 2y + 3\lambda(y^2 - ax) = 0, \qquad x^3 + y^3 - 3axy = 0$$

or
$$2x \cdot 3(y^2 - ax) - 2y \cdot 3(x^2 - ay) = 0, \qquad x^3 + y^3 - 3axy = 0$$

are the conditions in the two cases. These equations may be solved for (0, 0), $(1\frac{1}{2}a, 1\frac{1}{2}a)$, and some imaginary values. The value (0, 0) is singular and λ cannot be determined, but the point is evidently a minimum of $x^2 + y^2$ by inspection. The point $(1\frac{1}{2}a, 1\frac{1}{2}a)$ gives $\lambda = -1\frac{1}{3}a$. That the point is a (relative constrained) maximum of $x^2 + y^2$ is also seen by inspection. There is no need to examine d^2f. In most practical problems the examination of the conditions of the second order may be waived. This example is one which may be treated in polar coördinates by the ordinary methods; but it is noteworthy that if it could not be treated that way, the method of solution by eliminating one of the variables by solving the cubic $F(x, y) = 0$ would be unavailable and the methods of constrained maxima would be required.

EXERCISES

1. By total differentiation and division obtain dy/dx in these cases. Do not substitute in (1), but use the method by which it was derived.

(α) $ax^2 + 2bxy + cy^2 - 1 = 0$, ($\beta$) $x^4 + y^4 = 4a^2xy$, (γ) $(\cos x)^y - (\sin y)^x = 0$,
(δ) $(x^2 + y^2)^2 = a^2(x^2 - y^2)$, ($\epsilon$) $e^x + e^y = 2xy$, (ζ) $x^{-2}y^{-2} = \tan^{-1}xy$.

2. Obtain the second derivative d^2y/dx^2 in Ex. 1 (α), (β), (ϵ), (ζ) by differentiating the value of dy/dx obtained above. Compare with use of (3).

3. Prove $\dfrac{d^2y}{dx^2} = -\dfrac{F_y'^2 F_{xx}'' - 2F_x'F_y'F_{xy}'' + F_x'^2 F_{yy}''}{F_y'^3}$.

4. Find the radius of curvature of these curves:

(α) $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$, $R = 3(axy)^{\frac{1}{3}}$, (β) $x^{\frac{1}{2}} + y^{\frac{1}{2}} = a^{\frac{1}{2}}$, $R = 2\sqrt{(x+y)^3/a}$,

(γ) $b^2x^2 + a^2y^2 = a^2b^2$, (δ) $xy^2 = a^2(a - x)$, (ϵ) $(ax)^2 + (by)^{\frac{2}{3}} = 1$.

5. Find y', y'', y''' in case $x^3 + y^3 - 3axy = 0$.

6. Extend equations (3) to obtain y''' and reduce by Ex. 3.

7. Find tangents parallel to the x-axis for $(x^2 + y^2)^2 = 2a^2(x^2 - y^2)$.

8. Find tangents parallel to the y-axis for $(x^2 + y^2 + ax)^2 = a^2(x^2 + y^2)$.

9. If $b^2 < ac$ in $ax^2 + 2bxy + cy^2 + fx + gy + h = 0$, circumscribe about the curve a rectangle parallel to the axes. Check algebraically.

10. Sketch $x^3 + y^3 = x^2y^2 + \frac{1}{2}(x^2 + y^2)$ near the singular point (1, 1).

11. Find the singular points and discuss the curves near them:

(α) $x^3 + y^3 = 3axy$, (β) $(x^2 + y^2)^2 = 2a^2(x^2 - y^2)$,

(γ) $x^4 + y^4 = 2(x - y)^2$, (δ) $y^5 + 2xy^2 = x^2 + y^4$.

12. Make these functions maxima or minima subject to the given conditions. Discuss the work both with and without a multiplier:

(α) $\dfrac{a}{u\cos x} + \dfrac{b}{v\cos y}$, $a\tan x + b\tan y = c$. *Ans.* $\dfrac{\sin x}{\sin y} = \dfrac{u}{v}$.

(β) $x^2 + y^2$, $ax^2 + 2bxy + cy^2 = f$. Find axes of conic.

(γ) Find the shortest distance from a point to a line (in a plane).

13. Write the second and third total differentials of $F(x, y) = 0$ and compare with (3) and Ex. 5. Try this method of calculating in Ex. 2.

14. Show that $F_x'dx + F_y'dy = 0$ does and should give the tangent line to $F(x, y) = 0$ at the points (x, y) if $dx = \xi - x$ and $dy = \eta - y$, where ξ, η are the coördinates of points other than (x, y) on the tangent line. Why is the equation inapplicable at singular points of the curve?

59. More general cases of implicit functions. The problem of implicit functions may be generalized in two ways. In the first place a greater number of variables may occur in the function, as

$$F(x, y, z) = 0, \qquad F(x, y, z, \cdots, u) = 0;$$

and the question may be to solve the equation for one of the variables in terms of the others and to determine the partial derivatives of the chosen dependent variable. In the second place there may be several equations connecting the variables and it may be required to solve the equations for some of the variables in terms of the others and to determine the partial derivatives of the chosen dependent variables

with respect to the independent variables. In both cases the formal differentiation and attempted formal solution of the equations for the derivatives will indicate the results and the theorem under which the solution is proper.

Consider the case $F(x, y, z) = 0$ and form the differential.

$$dF(x, y, z) = F'_x dx + F'_y dy + F'_z dz = 0. \tag{8}$$

If z is to be the dependent variable, the partial derivative of z by x is found by setting $dy = 0$ so that y is constant. Thus

$$\frac{\partial z}{\partial x} = \left(\frac{dz}{dx}\right)_y = -\frac{F'_x}{F'_z} \quad \text{and} \quad \frac{\partial z}{\partial y} = \left(\frac{dz}{dy}\right)_x = -\frac{F'_y}{F'_z} \tag{9}$$

are obtained by ordinary division after setting $dy = 0$ and $dx = 0$ respectively. If this division is to be legitimate, F'_z must not vanish at the point considered. The immediate suggestion is the theorem: If, when real values (x_0, y_0) are chosen and a real value z_0 is obtained from $F(z, x_0, y_0) = 0$ by solution, the function $F(x, y, z)$ regarded as a function of three independent variables (x, y, z) is continuous at and near (x_0, y_0, z_0) and has continuous first partial derivatives and $F'_z(x_0, y_0, z_0) \neq 0$, then $F(x, y, z) = 0$ may be solved uniquely for $z = \phi(x, y)$ and $\phi(x, y)$ will be continuous and have partial derivatives (9) for values of (x, y) sufficiently near to (x_0, y_0).

The theorem is again proved by the Law of the Mean, and in a similar manner.

$$F(x, y, z) - F(x_0, y_0, z_0) = F(x, y, z) = (hF'_x + kF'_y + lF'_z)_{x_0+\theta h,\, y_0+\theta k,\, z_0+\theta l}.$$

As F'_x, F'_y, F'_z are continuous and $F'_z(x_0, y_0, z_0) \neq 0$, it is possible to take δ so small that, when $|h| < \delta$, $|k| < \delta$, $|l| < \delta$, the derivative $|F'_z| > m$ and $|F'_x| < \mu$, $|F'_y| < \mu$. Now it is desired so to restrict h, k that $\pm\delta F'_z$ shall determine the sign of the parenthesis. Let

$$|x - x_0| < \tfrac{1}{2} m\delta/\mu, \qquad |y - y_0| < \tfrac{1}{2} m\delta/\mu, \quad \text{then} \quad |hF'_x + kF'_y| < m\delta$$

and the signs of the parenthesis for $(x, y, z_0 + \delta)$ and $(x, y, z_0 - \delta)$ will be opposite since $|F'_z| > m$. Hence if (x, y) be held fixed, there is one and only one value of z for which the parenthesis vanishes between $z_0 + \delta$ and $z_0 - \delta$. Thus z is defined as a single valued function of (x, y) for sufficiently small values of $h = x - x_0$, $k = y - y_0$.

Also $$\frac{l}{h} = -\frac{F'_x(x_0 + \theta h, y_0 + \theta k, z_0 + \theta l)}{F'_z(x_0 + \theta h, y_0 + \theta k, z_0 + \theta l)}, \qquad \frac{l}{k} = -\frac{F'_y(\cdots)}{F'_z(\cdots)}$$

when k and h respectively are assigned the values 0. The limits exist when $h \doteq 0$ or $k \doteq 0$. But in the first case $l = \Delta z = \Delta_x z$ is the increment of z when x alone varies, and in the second case $l = \Delta z = \Delta_y z$. The limits are therefore the desired partial derivatives of z by x and y. The proof for any number of variables would be similar.

If none of the derivatives F'_x, F'_y, F'_z vanish, the equation $F(x, y, z) = 0$ may be solved for any one of the variables, and formulas like (9) will express the partial derivatives. It then appears that

$$\left(\frac{dz}{dx}\right)_y\left(\frac{dx}{dz}\right)_y = \frac{\partial z}{\partial x}\frac{\partial x}{\partial z} = \frac{F'_x}{F'_z}\frac{F'_z}{F'_x} = 1, \tag{10}$$

and
$$\left(\frac{dz}{dx}\right)_y\left(\frac{dx}{dy}\right)_z\left(\frac{dy}{dz}\right)_x = \frac{\partial z}{\partial x}\frac{\partial x}{\partial y}\frac{\partial y}{\partial z} = -1 \tag{11}$$

in like manner. The first equation is in this case identical with (4) of § 2 because if y is constant the relation $F(x, y, z) = 0$ reduces to $G(x, z) = 0$. The second equation is new. By virtue of (10) and similar relations, the derivatives in (11) may be inverted and transformed to the right side of the equation. As it is assumed in thermodynamics that the pressure, volume, and temperature of a given simple substance are connected by an equation $F(p, v, T) = 0$, called the characteristic equation of the substance, a relation between different thermodynamic magnitudes is furnished by (11).

60. In the next place suppose there are two equations

$$F(x, y, u, v) = 0, \qquad G(x, y, u, v) = 0 \tag{12}$$

between four variables. Let each equation be differentiated.

$$dF = 0 = F'_x dx + F'_y dy + F'_u du + F'_v dv,$$
$$dG = 0 = G'_x dx + G'_y dy + G'_u du + G'_v dv. \tag{13}$$

If it be desired to consider u, v as the dependent variables and x, y as independent, it would be natural to solve these equations for the differentials du and dv in terms of dx and dy; for example,

$$du = -\frac{(F'_x G'_v - F'_v G'_x)\,dx + (F'_y G'_v - F'_v G'_y)\,dy}{F'_u G'_v - F'_v G'_u}. \tag{13'}$$

The differential dv would have a different numerator but the same denominator. The solution requires $F'_u G'_v - F'_v G'_u \neq 0$. This suggests the desired theorem: If (u_0, v_0) are solutions of $F = 0$, $G = 0$ corresponding to (x_0, y_0) and if $F'_u G'_v - F'_v G'_u$ does not vanish for the values (x_0, y_0, u_0, v_0), the equations $F = 0$, $G = 0$ may be solved for $u = \phi(x, y)$, $v = \psi(x, y)$ and the solution is unique and valid for (x, y) sufficiently near (x_0, y_0) — it being assumed that F and G regarded as functions in four variables are continuous and have continuous first partial derivatives at and near (x_0, y_0, u_0, v_0); moreover, the total differentials du, dv are given by (13') and a similar equation.

The proof of this theorem may be deferred (§ 64). Some observations should be made. The equations (13) may be solved for any two variables in terms of the other two. The partial derivatives

$$\frac{\partial u(x, y)}{\partial x}, \qquad \frac{\partial u(x, v)}{\partial x}, \qquad \frac{\partial x(u, v)}{\partial u}, \qquad \frac{\partial x(u, y)}{\partial u} \tag{14}$$

of u by x or of x by u will naturally depend on whether the solution for u is in terms of (x, y) or of (x, v), and the solution for x is in (u, v) or (u, y). Moreover, it must not be assumed that $\partial u/\partial x$ and $\partial x/\partial u$ are reciprocals no matter which meaning is attached to each. In obtaining relations between the derivatives analogous to (10), (11), the values of the derivatives in terms of the derivatives of F and G may be found or the equations (12) may first be considered as solved.

Thus if $\qquad u = \phi(x, y), \qquad du = \phi_x' dx + \phi_y' dy,$

$\qquad v = \psi(x, y), \qquad dv = \psi_x' dx + \psi_y' dy.$

Then
$$dx = \frac{\psi_y' du - \phi_y' dv}{\phi_x'\psi_y' - \phi_y'\psi_x'}, \qquad dy = \frac{-\psi_x' du + \phi_x' dv}{\phi_x'\psi_y' - \phi_y'\psi_x'}$$

and
$$\frac{\partial x}{\partial u} = \frac{\psi_y'}{\phi_x'\psi_y' - \phi_y'\psi_x'}, \qquad \frac{\partial x}{\partial v} = \frac{-\phi_y'}{\phi_x'\psi_y' - \phi_y'\psi_x'}, \text{ etc.}$$

Hence
$$\frac{\partial u}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial v}{\partial x}\frac{\partial x}{\partial v} = 1, \tag{15}$$

as may be seen by direct substitution. Here u, v are expressed in terms of x, y for the derivatives u_x', v_x'; and x, y are considered as expressed in terms of u, v for the derivatives x_u', x_v'.

61. The questions of free or constrained maxima and minima, at any rate in so far as the determination of the conditions of the first order is concerned, may now be treated. If $F(x, y, z) = 0$ is given and the maxima and minima of z as a function of (x, y) are wanted,

$$F_x'(x, y, z) = 0, \qquad F_y'(x, y, z) = 0, \qquad F(x, y, z) = 0 \tag{16}$$

are three equations which may be solved for x, y, z. If for any of these solutions the derivative F_z' does not vanish, the surface $z = \phi(x, y)$ has at that point a tangent plane parallel to $z = 0$ and there is a maximum, minimum, or minimax. To distinguish between the possibilities further investigation must be made if necessary; the details of such an investigation will not be outlined for the reason that special methods are usually available. The conditions for an extreme of u as a function of (x, y) defined implicitly by the equations (13′) are seen to be

$$F_x'G_v' - F_v'G_x' = 0, \quad F_y'G_v' - F_v'G_y' = 0, \quad F = 0, \quad G = 0. \tag{17}$$

The four equations may be solved for x, y, u, v or merely for x, y.

Suppose that the maxima, minima, and minimax of $u = f(x, y, z)$ subject either to one equation $F(x, y, z) = 0$ or two equations $F(x, y, z) = 0$, $G(x, y, z) = 0$ of constraint are desired. Note that if only one equation of constraint is imposed, the function $u = f(x, y, z)$ becomes a function of two variables; whereas if two equations are imposed, the function u really contains only one variable and the question of a minimax does not arise. The *method of multipliers* is again employed. Consider

$$\Phi(x, y, z) = f + \lambda F \quad \text{or} \quad \Phi = f + \lambda F + \mu G \tag{18}$$

as the case may be. The conditions for a free extreme of Φ are

$$\Phi'_x = 0, \qquad \Phi'_y = 0, \qquad \Phi'_z = 0. \tag{19}$$

These three equations may be solved for the coördinates x, y, z which will then be expressed as functions of λ or of λ and μ according to the case. If then λ or λ and μ be determined so that (x, y, z) satisfy $F = 0$ or $F = 0$ and $G = 0$, the constrained extremes of $u = f(x, y, z)$ will be found except for the examination of the conditions of higher order.

As a problem in constrained maxima and minima let the axes of the section of an ellipsoid by a plane through the origin be determined. Form the function

$$\Phi = x^2 + y^2 + z^2 + \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1\right) + \mu(lx + my + nz)$$

by adding to $x^2 + y^2 + z^2$, which is to be made extreme, the equations of the ellipsoid and plane, which are the equations of constraint. Then apply (19). Hence

$$x + \lambda\frac{x}{a^2} + \frac{\mu}{2}l = 0, \qquad y + \lambda\frac{y}{b^2} + \frac{\mu}{2}m = 0, \qquad z + \lambda\frac{z}{c^2} + \frac{\mu}{2}n = 0$$

taken with the equations of ellipsoid and plane will determine x, y, z, λ, μ. If the equations are multiplied by x, y, z and reduced by the equations of plane and ellipsoid, the solution for λ is $\lambda = -r^2 = -(x^2 + y^2 + z^2)$. The three equations then become

$$x = \frac{1}{2}\frac{\mu l a^2}{r^2 - a^2}, \qquad y = \frac{1}{2}\frac{\mu m b^2}{r^2 - b^2}, \qquad z = \frac{1}{2}\frac{\mu n c^2}{r^2 - c^2}, \qquad \text{with} \quad lx + my + nz = 0.$$

Hence
$$\frac{l^2a^2}{r^2 - a^2} + \frac{m^2b^2}{r^2 - b^2} + \frac{n^2c^2}{r^2 - c^2} = 0 \quad \text{determines } r^2. \tag{20}$$

The two roots for r are the major and minor axes of the ellipse in which the plane cuts the ellipsoid. The substitution of x, y, z above in the ellipsoid determines

$$\frac{\mu^2}{4}\left[\left(\frac{al}{r^2 - a^2}\right)^2 + \left(\frac{bm}{r^2 - b^2}\right)^2 + \left(\frac{cn}{r^2 - c^2}\right)^2\right] \quad \text{since} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{21}$$

Now when (20) is solved for any particular root r and the value of μ is found by (21), the actual coördinates x, y, z of the extremities of the axes may be found.

EXERCISES

1. Obtain the partial derivatives of z by x and y directly from (8) and not by substitution in (9). Where does the solution fail?

$(\alpha)\ \dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}+\dfrac{z^2}{c^2}=1,\qquad (\beta)\ x+y+z=\dfrac{1}{xyz},$

$(\gamma)\ (x^2+y^2+z^2)^2=a^2x^2+b^2y^2+c^2z^2,\qquad (\delta)\ xyz=c.$

2. Find the second derivatives in Ex. 1 (α), (β), (δ) by repeated differentiation.

3. State and prove the theorem on the solution of $F(x, y, z, u)=0$.

4. Show that the product $\alpha_p E_T$ of the coefficient of expansion by the modulus of elasticity (§ 52) is equal to the rate of rise of pressure with the temperature if the volume is constant.

5. Establish the proportion $E_S : E_T = C_p : C_v$ (see § 52).

6. If $F(x, y, z, u)=0$, show $\dfrac{\partial u}{\partial x}\dfrac{\partial x}{\partial y}\dfrac{\partial y}{\partial z}\dfrac{\partial z}{\partial u}=1,\quad \dfrac{\partial u}{\partial x}\dfrac{\partial x}{\partial u}=1.$

7. Write the equations of tangent plane and normal line to $F(x, y, z)=0$ and find the tangent planes and normal lines to Ex. 1 (β), (δ) at $x=1$, $y=1$.

8. Find, by using (13), the indicated derivatives on the assumption that either x, y or u, v are dependent and the other pair independent:

$(\alpha)\ u^5+v^5+x^5-3y=0,\qquad u^3+v^3+y^3+3x=0,\qquad u'_x,\ u'_y,\ u''_{xy},\ v''_{xx}$

$(\beta)\ x+y+u+v=a,\qquad x^2+y^2+u^2+v^2=b,\qquad x'_u,\ u'_x,\ v'_y,\ v''_{xy}$

(γ) Find dy in both cases if x, v are independent variables.

9. Prove $\dfrac{\partial u}{\partial x}\dfrac{\partial y}{\partial u}+\dfrac{\partial v}{\partial x}\dfrac{\partial y}{\partial v}=0$ if $F(x, y, u, v)=0$, $G(x, y, u, v)=0$.

10. Find du and the derivatives u'_x, u'_y, u'_z in case

$$x^2+y^2+z^2=uv,\qquad xy=u^2+v^2+w^2,\qquad xyz=uvw.$$

11. If $F(x, y, z)=0$, $G(x, y, z)=0$ define a curve, show that

$$\frac{x-x_0}{(F'_yG'_z-F'_zG'_y)_0}=\frac{y-y_0}{(F'_zG'_x-F'_xG'_z)_0}=\frac{z-z_0}{(F'_xG'_y-F'_yG'_x)_0}$$

is the tangent line to the curve at (x_0, y_0, z_0). Write the normal plane.

12. Formulate the problem of implicit functions occurring in Ex. 11.

13. Find the perpendicular distance from a point to a plane.

14. The sum of three positive numbers is $x+y+z=N$, where N is given. Determine x, y, z so that the product $x^py^qz^r$ shall be maximum if p, q, r are given.

Ans. $x:y:z:N=p:q:r:(p+q+r)$.

15. The sum of three positive numbers and the sum of their squares are both given. Make the product a maximum or minimum.

16. The surface $(x^2+y^2+z^2)^2=ax^2+by^2+cz^2$ is cut by the plane $lx+my+nz=0$. Find the maximum or minimum radius of the section. *Ans.* $\sum\dfrac{l^2}{r^2-a}=0$.

17. In case $F(x, y, u, v) = 0$, $G(x, y, u, v) = 0$ consider the differentials

$$dv = \frac{\partial v}{\partial x}dx + \frac{\partial v}{\partial y}dy, \qquad dx = \frac{\partial x}{\partial u}du + \frac{\partial x}{\partial v}dv, \qquad dy = \frac{\partial y}{\partial u}du + \frac{\partial y}{\partial v}dv.$$

Substitute in the first from the last two and obtain relations like (15) and Ex. 9.

18. If $f(x, y, z)$ is to be maximum or minimum subject to the constraint $F(x, y, z) = 0$, show that the conditions are that $dx : dy : dz = 0 : 0 : 0$ are indeterminate when their solution is attempted from

$$f'_x dx + f'_y dy + f'_z dz = 0 \quad \text{and} \quad F'_x dx + F'_y dy + F'_z dz = 0.$$

From what geometrical considerations should this be obvious? Discuss in connection with the problem of inscribing the maximum rectangular parallelepiped in the ellipsoid. These equations,

$$dx : dy : dz = f'_y F'_z - f'_z F'_y : f'_z F'_x - f'_x F'_z : f'_x F'_y - f'_y F'_x = 0 : 0 : 0,$$

may sometimes be used to advantage for such problems.

19. Given the curve $F(x, y, z) = 0$, $G(x, y, z) = 0$. Discuss the conditions for the highest or lowest points, or more generally the points where the tangent is parallel to $z = 0$, by treating $u = f(x, y, z) = z$ as a maximum or minimum subject to the two constraining equations $F = 0$, $G = 0$. Show that the condition $F'_x G'_y = F'_y G'_x$ which is thus obtained is equivalent to setting $dz = 0$ in

$$F'_x dx + F'_y dy + F'_z dz = 0 \quad \text{and} \quad G'_x dx + G'_y dy + G'_z dz = 0.$$

20. Find the highest and lowest points of these curves:

(α) $x^2 + y^2 = z^2 + 1$, $x + y + 2z = 0$, $\qquad (\beta)$ $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$, $lx + my + nz = 0$.

21. Show that $F'_x dx + F'_y dy + F'_z dz = 0$, with $dx = \xi - x$, $dy = \eta - y$, $dz = \zeta - z$, is the tangent plane to the surface $F(x, y, z) = 0$ at (x, y, z). Apply to Ex. 1.

22. Given $F(x, y, u, v) = 0$, $G(x, y, u, v) = 0$. Obtain the equations

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial x} = 0, \qquad \frac{\partial F}{\partial y} + \frac{\partial F}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial y} = 0,$$

$$\frac{\partial G}{\partial x} + \frac{\partial G}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial G}{\partial v}\frac{\partial v}{\partial x} = 0, \qquad \frac{\partial G}{\partial y} + \frac{\partial G}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial G}{\partial v}\frac{\partial v}{\partial y} = 0,$$

and explain their significance as a sort of partial-total differentiation of $F = 0$ and $G = 0$. Find u'_x from them and compare with (13′). Write similar equations where x, y are considered as functions of (u, v). Hence prove, and compare with (15) and Ex. 9,

$$\frac{\partial u}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial v}{\partial y}\frac{\partial y}{\partial v} = 1, \qquad \frac{\partial u}{\partial y}\frac{\partial x}{\partial u} + \frac{\partial v}{\partial y}\frac{\partial x}{\partial v} = 0.$$

23. Show that the differentiation with respect to x and y of the four equations under Ex. 22 leads to eight equations from which the eight derivatives

$$\frac{\partial^2 u}{\partial x^2}, \quad \frac{\partial^2 u}{\partial x \partial y}, \quad \frac{\partial^2 u}{\partial y \partial x}, \quad \frac{\partial^2 u}{\partial y^2}, \quad \frac{\partial^2 v}{\partial x^2}, \quad \cdots, \quad \frac{\partial^2 v}{\partial y^2}$$

may be obtained. Show thus that formally $u''_{xy} = u''_{yx}$.

62. Functional determinants or Jacobians. Let two functions

$$u = \phi(x, y), \qquad v = \psi(x, y) \tag{22}$$

of two independent variables be given. The continuity of the functions and of their first derivatives is assumed throughout this discussion and will not be mentioned again. Suppose that there were a relation $F(u, v) = 0$ or $F(\phi, \psi) = 0$ between the functions. Then

$$F(\phi, \psi) = 0, \qquad F'_u\phi'_x + F'_v\psi'_x = 0, \qquad F'_u\phi'_y + F'_v\psi'_y = 0. \tag{23}$$

The last two equations arise on differentiating the first with respect to x and y. The elimination of F'_u and F'_v from these gives

$$\phi'_x\psi'_y - \phi'_y\psi'_x = \begin{vmatrix} \phi'_x & \psi'_x \\ \phi'_y & \psi'_y \end{vmatrix} = \frac{\partial(u, v)}{\partial(x, y)} = J\left(\frac{u, v}{x, y}\right) = 0. \tag{24}$$

The determinant is merely another way of writing the first expression; the next form is the customary short way of writing the determinant and denotes that the elements of the determinant are the first derivatives of u and v with respect to x and y. This determinant is called the *functional determinant* or *Jacobian* of the functions u, v or ϕ, ψ with respect to the variables x, y and is denoted by J. It is seen that: *If there is a functional relation $F(\phi, \psi) = 0$ between two functions, the Jacobian of the functions vanishes identically*, that is, vanishes for all values of the variables (x, y) under consideration.

Conversely, *if the Jacobian vanishes identically over a two-dimensional region for (x, y), the functions are connected by a functional relation.* For, the functions u, v may be assumed not to reduce to mere constants and hence there may be assumed to be points for which at least one of the partial derivatives ϕ'_x, ϕ'_y, ψ'_x, ψ'_y does not vanish. Let ϕ'_x be the derivative which does not vanish at some particular point of the region. Then $u = \phi(x, y)$ may be solved as $x = \chi(u, y)$ in the vicinity of that point and the result may be substituted in v.

$$v = \psi(\chi, y), \qquad \frac{\partial v}{\partial y} = \psi'_x\frac{\partial \chi}{\partial y} + \psi'_y = \psi'_x\frac{\partial x}{\partial y} + \psi'_y.$$

But $$\frac{\partial x}{\partial y} = -\frac{\partial u}{\partial y}\frac{\partial x}{\partial u} \quad \text{and} \quad \frac{\partial v}{\partial y} = \frac{1}{\phi'_x}(\phi'_x\psi'_y - \psi'_x\phi'_y) \tag{24'}$$

by (11) and substitution. Thus $\partial v/\partial y = J/\phi'_x$; and if $J = 0$, then $\partial v/\partial y = 0$. This relation holds at least throughout the region for which $\phi'_x \neq 0$, and for points in this region $\partial v/\partial y$ vanishes identically. Hence v does not depend on y but becomes a function of u alone. This establishes the fact that v and u are functionally connected.

These considerations may be extended to other cases. Let

$$u = \phi(x, y, z), \qquad v = \psi(x, y, z), \qquad w = \chi(x, y, z). \qquad (25)$$

If there is a functional relation $F(u, v, w) = 0$, differentiate it.

$$\begin{aligned} F'_u\phi'_x + F'_v\psi'_x + F'_w\chi'_x &= 0, \\ F'_u\phi'_y + F'_v\psi'_y + F'_w\chi'_y &= 0, \\ F'_u\phi'_z + F'_v\psi'_z + F'_w\chi'_z &= 0, \end{aligned} \qquad \begin{vmatrix} \phi'_x & \psi'_x & \chi'_x \\ \phi'_y & \psi'_y & \chi'_y \\ \phi'_z & \psi'_z & \chi'_z \end{vmatrix} = 0, \qquad (26)$$

or

$$\frac{\partial(\phi, \psi, \chi)}{\partial(x, y, z)} = \frac{\partial(u, v, w)}{\partial(x, y, z)} = J = 0.$$

The result is obtained by eliminating F'_u, F'_v, F'_w from the three equations. The assumption is made, here as above, that F'_u, F'_v, F'_w do not all vanish; for if they did, the three equations would not imply $J = 0$. On the other hand their vanishing would imply that F did not contain u, v, w, — as it must if there is really a relation between them. And now conversely it may be shown that if J vanishes identically, there is a functional relation between u, v, w. Hence again *the necessary and sufficient conditions that the three functions (25) be functionally connected is that their Jacobian vanish.*

The proof of the converse part is about as before. It may be assumed that at least one of the derivatives of u, v, w or ϕ, ψ, χ by x, y, z does not vanish. Let $\phi'_x \neq 0$ be that derivative. Then $u = \phi(x, y, z)$ may be solved as $x = \omega(u, y, z)$ and the result may be substituted in v and w as

$$v = \psi(x, y, z) = \psi(\omega, y, z), \qquad w = \chi(x, y, z) = \chi(\omega, y, z).$$

Next the Jacobian of v and w relative to y and z may be written as

$$\begin{vmatrix} \dfrac{\partial v}{\partial y} & \dfrac{\partial w}{\partial y} \\ \dfrac{\partial v}{\partial z} & \dfrac{\partial w}{\partial z} \end{vmatrix} = \begin{vmatrix} \psi'_x\dfrac{\partial x}{\partial y} + \psi'_y & \chi'_x\dfrac{\partial x}{\partial y} + \chi'_y \\ \psi'_x\dfrac{\partial x}{\partial z} + \psi'_z & \chi'_x\dfrac{\partial x}{\partial z} + \chi'_z \end{vmatrix}$$

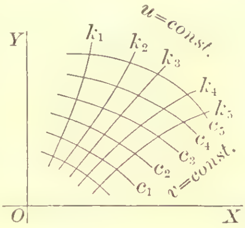
$$= \begin{vmatrix} \psi'_y & \chi'_y \\ \psi'_z & \chi'_z \end{vmatrix} + \psi'_x \begin{vmatrix} -\phi'_y/\phi'_x & \chi'_y \\ -\phi'_z/\phi'_x & \chi'_z \end{vmatrix} + \chi'_x \begin{vmatrix} \psi'_y & -\phi'_y/\phi'_x \\ \psi'_z & -\phi'_z/\phi'_x \end{vmatrix}$$

$$= \frac{1}{\phi'_x}\left[\phi'_x \begin{vmatrix} \psi'_y & \chi'_y \\ \psi'_z & \chi'_z \end{vmatrix} + \psi'_x \begin{vmatrix} \chi'_y & \phi'_y \\ \chi'_z & \phi'_z \end{vmatrix} + \chi'_x \begin{vmatrix} \phi'_y & \psi'_y \\ \phi'_z & \psi'_z \end{vmatrix}\right] = \frac{J}{\phi'_x}.$$

As J vanishes identically, the Jacobian of v and w expressed as functions of y, z, also vanishes. Hence by the case previously discussed there is a functional relation $F(v, w) = 0$ independent of y, z; and as v, w now contain u, this relation may be considered as a functional relation between u, v, w.

63. If in (22) the variables u, v be assigned constant values, the equations define two curves, and if u, v be assigned a series of such values, the equations (22) define a network of curves in some part of the

xy-plane. If there is a functional relation $u = F(v)$, that is, if the Jacobian vanishes identically, a constant value of v implies a constant value of u and hence the locus for which v is constant is also a locus for which u is constant; the set of v-curves coincides with the set of u-curves and no true network is formed. This case is uninteresting. Let it be assumed that the Jacobian does not vanish identically and even that it does not vanish for any point (x, y) of a certain region of the xy-plane. The indications of § 60 are that the equations (22) may then be solved for x, y in terms of u, v at any point of the region and that there is a pair of the curves through each point. It is then proper to consider (u, v) as the coördinates of the points in the region. To any point there correspond not only the rectangular coördinates (x, y) but also the *curvilinear coördinates* (u, v).



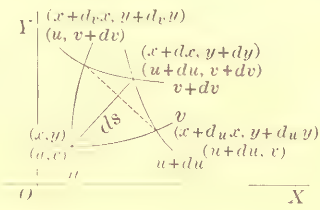
The equations connecting the rectangular and curvilinear coördinates may be taken in either of the two forms

$$u = \phi(x, y), \qquad v = \psi(x, y) \quad \text{or} \quad x = f(u, v), \qquad y = g(u, v), \tag{22'}$$

each of which are the solutions of the other. The Jacobians

$$J\left(\frac{u, v}{x, y}\right) \cdot J\left(\frac{x, y}{u, v}\right) = 1 \tag{27}$$

are reciprocal each to each; and this relation may be regarded as the analogy of the relation (4) of § 2 for the case of the function $y = \phi(x)$ and the solution $x = f(y) = \phi^{-1}(y)$ in the case of a single variable. The *differential of arc* is



$$ds^2 = dx^2 + dy^2 = E du^2 + 2\,F du dv + G dv^2, \tag{28}$$

$$E = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2, \qquad F = \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v}, \qquad G = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2.$$

The *differential of area* included between two neighboring u-curves and two neighboring v-curves may be written in the form

$$dA = J\left(\frac{x, y}{u, v}\right) du dv = du dv \div J\left(\frac{u, v}{x, y}\right). \tag{29}$$

These statements will now be proved in detail.

To prove (27) write out the Jacobians at length and reduce the result.

$$J\left(\frac{u,\ v}{x,\ y}\right)J\left(\frac{x,\ y}{u,\ v}\right)=\begin{vmatrix}\frac{\partial u}{\partial x} & \frac{\partial v}{\partial x}\\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y}\end{vmatrix}\cdot\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial y}{\partial u}\\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v}\end{vmatrix}$$

$$=\begin{vmatrix}\frac{\partial u}{\partial x}\frac{\partial x}{\partial u}+\frac{\partial v}{\partial x}\frac{\partial x}{\partial v} & \frac{\partial u}{\partial x}\frac{\partial y}{\partial u}+\frac{\partial v}{\partial x}\frac{\partial y}{\partial v}\\ \frac{\partial u}{\partial y}\frac{\partial x}{\partial u}+\frac{\partial v}{\partial y}\frac{\partial x}{\partial v} & \frac{\partial u}{\partial y}\frac{\partial y}{\partial u}+\frac{\partial v}{\partial y}\frac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}1 & 0\\ 0 & 1\end{vmatrix}=1,$$

where the rule for multiplying determinants has been applied and the reduction has been made by (15), Ex. 9 above, and similar formulas. If the rule for multiplying determinants is unfamiliar, the Jacobians may be written and multiplied without that notation and the reduction may be made by the same formulas as before.

To establish the formula for the differential of arc it is only necessary to write the total differentials of dx and dy, to square and add, and then collect. To obtain the differential area between four adjacent curves consider the triangle determined by $(u,\ v)$, $(u+du,\ v)$, $(u,\ v+dv)$, which is half that area, and double the result. The determinantal form of the area of a triangle is the best to use.

$$dA=2\cdot\frac{1}{2}\begin{vmatrix}d_ux & d_uy\\ d_vx & d_vy\end{vmatrix}=\begin{vmatrix}\frac{\partial x}{\partial u}du & \frac{\partial y}{\partial u}du\\ \frac{\partial x}{\partial v}dv & \frac{\partial y}{\partial v}dv\end{vmatrix}=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial y}{\partial u}\\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v}\end{vmatrix}dudv.$$

The subscripts on the differentials indicate which variable changes; thus d_ux, d_uy are the coördinates of $(u+du,\ v)$ relative to $(u,\ v)$. This method is easily extended to determine the analogous quantities in three dimensions or more. It may be noticed that the triangle does not look as if it were half the area (except for infinitesimals of higher order) in the figure; but see Ex. 12 below.

It should be remarked that as the differential of area dA is usually considered positive when du and dv are positive, it is usually better to replace J in (29) by its absolute value. Instead of regarding $(u,\ v)$ as curvilinear coördinates in the xy-plane, it is possible to plot them in their own uv-plane and thus to establish by (22′) a *transformation* of the xy-plane over onto the uv-plane. A small area in the xy-plane then becomes a small area in the uv-plane. If $J>0$, the transformation is called direct; but if $J<0$, the transformation is called perverted. The significance of the distinction can be made clear only when the question of the signs of areas has been treated. The transformation is called *conformal* when elements of arc in the neighborhood of a point in the xy-plane are proportional to the elements of arc in the neighborhood of the corresponding point in the uv-plane, that is, when

$$ds^2=dx^2+dy^2=k(du^2+dv^2)=kd\sigma^2. \tag{30}$$

For in this case any little triangle will be transformed into a little triangle similar to it, and hence angles will be unchanged by the transformation. That the transformation be conformal requires that $F = 0$ and $E = G$. It is not necessary that $E = G = k$ be constants; the ratio of similitude may be different for different points.

64. There remains outstanding the proof that equations may be solved in the neighborhood of a point at which the Jacobian does not vanish. The fact was indicated in § 60 and used in § 63.

THEOREM. Let p equations in $n + p$ variables be given, say,

$$F_1(x_1, x_2, \cdots, x_{n+p}) = 0, \qquad F_2 = 0, \cdots, F_p = 0. \tag{31}$$

Let the p functions be soluble for $x_{10}, x_{20}, \cdots, x_{p0}$ when a particular set $x_{(p+1)0}, \cdots, x_{(n+p)0}$ of the other n variables are given. Let the functions and their first derivatives be continuous in all the $n + p$ variables in the neighborhood of $(x_{10}, x_{20}, \cdots, x_{(n+p)0})$. Let the Jacobian of the functions with respect to $x_1, x_2, \cdots, x_p$,

$$J\left(\frac{F_1, \cdots, F_p}{x_1, \cdots, x_p}\right) = \left|\begin{matrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_p}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial F_1}{\partial x_p} & \cdots & \frac{\partial F_p}{\partial x_p} \end{matrix}\right|_{x_{10}, \cdots, x_{(n+p)0}} \neq 0, \tag{32}$$

fail to vanish for the particular set mentioned. Then the p equations may be solved for the p variables $x_1, x_2, \cdots, x_p$, and the solutions will be continuous, unique, and differentiable with continuous first partial derivatives for all values of $x_{p+1}, \cdots, x_{n+p}$ sufficiently near to the values $x_{(p+1)0}, \cdots, x_{(n+p)0}$.

THEOREM. The necessary and sufficient condition that a functional relation exist between p functions of p variables is that the Jacobian of the functions with respect to the variables shall vanish identically, that is, for all values of the variables.

The proofs of these theorems will naturally be given by mathematical induction. Each of the theorems has been proved in the simplest cases and it remains only to show that the theorems are true for p functions in case they are for $p - 1$. Expand the determinant J.

$$J = J_1 \frac{\partial F_1}{\partial x_1} + J_2 \frac{\partial F_1}{\partial x_2} + \cdots + J_p \frac{\partial F_1}{\partial x_p}, \qquad J_1, \cdots, J_p, \text{ minors.}$$

For the first theorem $J \neq 0$ and hence at least one of the minors $J_1, \cdots, J_p$ must fail to vanish. Let that one be J_1, which is the Jacobian of $F_2, \cdots, F_p$ with respect to $x_2, \cdots, x_p$. By the assumption that the theorem holds for the case $p - 1$, these $p - 1$ equations may be solved for $x_2, \cdots, x_p$ in terms of the $n + 1$ variables x_1,

$x_{p+1}, \cdots, x_{n+p}$, and the results may be substituted in F_1. It remains to show that $F_1 = 0$ is soluble for x_1. Now

$$\frac{dF_1}{dx_1} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_1}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \cdots + \frac{\partial F_1}{\partial x_p}\frac{\partial x_p}{\partial x_1} = J/J_1 \neq 0. \tag{32'}$$

For the derivatives of $x_2, \cdots, x_p$ with respect to x_1 are obtained from the equations

$$0 = \frac{\partial F_2}{\partial x_1} + \frac{\partial F_2}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \cdots + \frac{\partial F_2}{\partial x_p}\frac{\partial x_p}{\partial x_1}, \quad \cdots, \quad 0 = \frac{\partial F_p}{\partial x_1} + \frac{\partial F_p}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \cdots + \frac{\partial F_p}{\partial x_p}\frac{\partial x_p}{\partial x_1}$$

resulting from the differentiation of $F_2 = 0, \cdots, F_p = 0$ with respect to x_1. The derivative $\partial x_i/\partial x_1$ is therefore merely J_i/J_1, and hence $dF_1/dx_1 = J/J_1$ and does not vanish. The equation therefore may be solved for x_1 in terms of $x_{p+1}, \cdots, x_{n+p}$, and this result may be substituted in the solutions above found for $x_2, \cdots, x_p$. Hence the equations have been solved for $x_1, x_2, \cdots, x_p$ in terms of $x_{p+1}, \cdots, x_{n+p}$ and the theorem is proved.

For the second theorem the procedure is analogous to that previously followed. If there is a relation $F(u_1, \cdots, u_p) = 0$ between the p functions

$$u_1 = \phi_1(x_1, \cdots, x_p), \cdots, \qquad u_p = \phi_p(x_1, \cdots, x_p),$$

differentiation with respect to $x_1, \cdots, x_p$ gives p equations from which the derivatives of F by $u_1, \cdots, u_p$ may be eliminated and $J\left(\frac{u_1, \cdots, u_p}{x_1, \cdots, x_p}\right) = 0$ becomes the condition desired. If conversely this Jacobian vanishes identically and it be assumed that one of the derivatives of u_i by x_j, say $\partial u_1/\partial x_1$, does not vanish, then the solution $x_1 = \omega(u_1, x_2, \cdots, x_p)$ may be effected and the result may be substituted in $u_2, \cdots, u_p$. The Jacobian of $u_2, \cdots, u_p$ with respect to $x_2, \cdots, x_p$ will then turn out to be $J \div \partial u_1/\partial x_1$ and will vanish because J vanishes. Now, however, only $p - 1$ functions are involved, and hence if the theorem is true for $p - 1$ functions it must be true for p functions.

EXERCISES

1. If $u = ax + by + c$ and $v = a'x + b'y + c'$ are functionally dependent, the lines $u = 0$ and $v = 0$ are parallel; and conversely.

2. Prove $x + y + z$, $xy + yz + zx$, $x^2 + y^2 + z^2$ functionally dependent.

3. If $u = ax + by + cz + d$, $v = a'x + b'y + c'z + d'$, $w = a''x + b''y + c''z + d''$ are functionally dependent, the planes $u = 0$, $v = 0$, $w = 0$ are parallel to a line.

4. In what senses are $\frac{\partial v}{\partial y}$ and ψ_y' of (24′) and $\frac{dF_1}{dx_1}$ and $\frac{\partial F_1}{\partial x_1}$ of (32′) partial or total derivatives? Are not the two sets completely analogous?

5. Given (25), suppose $\begin{vmatrix} \psi_y' & \chi_y' \\ \psi_z' & \chi_z' \end{vmatrix} \neq 0$. Solve $v = \psi$ and $w = \chi$ for y and z, substitute in $u = \phi$, and prove $\partial u/\partial x = J \div \begin{vmatrix} \psi_y' & \chi_y' \\ \psi_z' & \chi_z' \end{vmatrix}$.

6. If $u = u(x, y)$, $v = v(x, y)$, and $x = x(\xi, \eta)$, $y = y(\xi, \eta)$, prove

$$J\left(\frac{u, v}{x, y}\right) J\left(\frac{x, y}{\xi, \eta}\right) = J\left(\frac{u, v}{\xi, \eta}\right). \tag{27'}$$

State the extension to any number of variables. How may (27′) be used to prove (27)? Again state the extension to any number of variables.

7. Prove $dV = J\left(\frac{x,\ y,\ z}{u,\ v,\ w}\right) du dv dw = du dv dw \div J\left(\frac{u,\ v,\ w}{x,\ y,\ z}\right)$ is the element of volume in space with curvilinear coördinates $u, v, w =$ consts.

8. In what parts of the plane can $u = x^2 + y^2$, $v = xy$ not be used as curvilinear coördinates? Express ds^2 for these coördinates.

9. Prove that $2u = x^2 - y^2$, $v = xy$ is a conformal transformation.

10. Prove that $x = \dfrac{u}{u^2 + v^2}$, $y = \dfrac{v}{u^2 + v^2}$ is a conformal transformation.

11. Define conformal transformation in space. If the transformation

$$x = au + bv + cw, \qquad y = a'u + b'v + c'w, \qquad z = a''u + b''v + c''w$$

is conformal, is it orthogonal? See Ex. 10 (ζ), p. 100.

12. Show that the areas of the triangles whose vertices are

$$(u, v), (u + du, v), (u, v + dv) \quad \text{and} \quad (u + du, v + dv), (u + du, v), (u, v + dv)$$

are infinitesimals of the same order, as suggested in § 63.

13. Would the condition $F = 0$ in (28) mean that the set of curves $u =$ const. were perpendicular to the set $v =$ const.?

14. Express E, F, G in (28) in terms of the derivatives of u, v by x, y.

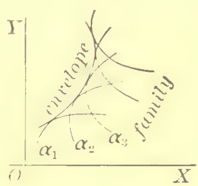
15. If $x = \phi(s, t)$, $y = \psi(s, t)$, $z = \chi(s, t)$ are the parametric equations of a surface (from which s, t could be eliminated to obtain the equation between x, y, z), show

$$\frac{\partial z}{\partial x} = J\left(\frac{\chi,\ \psi}{s,\ t}\right) \div J\left(\frac{\phi,\ \psi}{s,\ t}\right) \quad \text{and find} \quad \frac{\partial z}{\partial y}.$$

65. Envelopes of curves and surfaces. Let the equation $F(x, y, \alpha) = 0$ be considered as representing a family of curves where the different curves of the family are obtained by assigning different values to the parameter α. Such families are illustrated by

$$(x - \alpha)^2 + y^2 = 1 \quad \text{and} \quad \alpha x + y/\alpha = 1, \tag{33}$$

which are circles of unit radius centered on the x-axis and lines which cut off the area $\frac{1}{2}\alpha^2$ from the first quadrant. As α changes, the circles remain always tangent to the two lines $y = \pm 1$ and the point of tangency traces those lines. Again, as α changes, the lines (33) remain tangent to the hyperbola $xy = k$, owing to the property of the hyperbola that a tangent forms a triangle of constant area with the asymptotes. The lines $y = \pm 1$ are called the *envelope* of the system of circles and the hyperbola $xy = k$ the envelope of the set of lines. In general, *if there is a curve to which the curves of a family $F(x, y, \alpha) = 0$ are tangent and if the point of tangency describes that curve as α varies, the curve is called*



the envelope (or part of the envelope if there are several such curves) *of the family* $F(x, y, \alpha) = 0$. Thus any curve may be regarded as the envelope of its tangents or as the envelope of its circles of curvature.

To find the equations of the envelope note that by definition the enveloping curves of the family $F(x, y, \alpha) = 0$ are tangent to the envelope and that the point of tangency moves along the envelope as α varies. The equation of the envelope may therefore be written

$$x = \phi(\alpha), \qquad y = \psi(\alpha) \quad \text{with} \quad F(\phi, \psi, \alpha) = 0, \tag{34}$$

where the first equations express the dependence of the points on the envelope upon the parameter α and the last equation states that each point of the envelope lies also on some curve of the family $F(x, y, \alpha) = 0$. Differentiate (34) with respect to α. Then

$$F'_x\phi'(\alpha) + F'_y\psi'(\alpha) + F'_\alpha = 0. \tag{35}$$

Now if the point of contact of the envelope with the curve $F = 0$ is an ordinary point of that curve, the tangent to the curve is

$$F'_x(x - x_0) + F'_y(y - y_0) = 0; \quad \text{and} \quad F'_x\phi' + F'_y\psi' = 0,$$

since the tangent direction $dy : dx = \psi' : \phi'$ along the envelope is by definition identical with that along the enveloping curve; and if the point of contact is a singular point for the enveloping curve, $F'_x = F'_y = 0$. Hence in either case $F'_\alpha = 0$.

Thus *for points on the envelope the two equations*

$$F(x, y, \alpha) = 0, \qquad F'_\alpha(x, y, \alpha) = 0 \tag{36}$$

are satisfied and the equation of the envelope of the family $F = 0$ *may be found by solving* (36) *to find the parametric equations* $x = \phi(\alpha)$, $y = \psi(\alpha)$ *of the envelope or by eliminating* α *between* (36) *to find the equation of the envelope in the form* $\Phi(x, y) = 0$. It should be remarked that the locus found by this process may contain other curves than the envelope. For instance if the curves of the family $F = 0$ have singular points and if $x = \phi(\alpha)$, $y = \psi(\alpha)$ be the locus of the singular points as α varies, equations (34), (35) still hold and hence (36) also. The rule for finding the envelope therefore finds also the locus of singular points. Other extraneous factors may also be introduced in performing the elimination. It is therefore important to test graphically or analytically the solution obtained by applying the rule.

As a first example let the envelope of $(x - \alpha)^2 + y^2 = 1$ be found.

$$F(x, y, \alpha) = (x - \alpha)^2 + y^2 - 1 = 0, \qquad F'_\alpha = -2(x - \alpha) = 0.$$

The elimination of α from these equations gives $y^2 - 1 = 0$ and the solution for α gives $x = \alpha$, $y = \pm 1$. The loci indicated as envelopes are $y = \pm 1$. It is

geometrically evident that these are really envelopes and not extraneous factors. But as a second example consider $\alpha x + y/\alpha = 1$. Here

$$F(x, y, \alpha) = \alpha x + y/\alpha - 1 = 0, \qquad F'_\alpha = x - y/\alpha^2 = 0.$$

The solution is $y = \alpha/2$, $x = 1/2\alpha$, which gives $xy = \frac{1}{4}$. This is the envelope; it could not be a locus of singular points of $F = 0$ as there are none. Suppose the elimination of α be made by Sylvester's method as

$$\begin{array}{rrl} -y/\alpha^2 & +0/\alpha & +x+0\alpha=0 \\ 0/\alpha^2 & -y/\alpha & +0+x\alpha=0 \\ y/\alpha^2 & -1/\alpha & +x+0\alpha=0 \\ 0/\alpha^2 & +y/\alpha & -1+x\alpha=0 \end{array} \quad \text{and} \quad \begin{vmatrix} -y & 0 & x & 0 \\ 0 & -y & 0 & x \\ y & -1 & x & 0 \\ 0 & y & -1 & x \end{vmatrix} = 0;$$

the reduction of the determinant gives $xy(4xy - 1) = 0$ as the eliminant, and contains not only the envelope $4xy = 1$, but the factors $x = 0$ and $y = 0$ which are obviously extraneous.

As a third problem find the envelope of a line of which the length intercepted between the axes is constant. The necessary equations are

$$\frac{x}{\alpha} + \frac{y}{\beta} = 1, \qquad \alpha^2 + \beta^2 = K^2, \qquad \frac{x}{\alpha^2}d\alpha + \frac{y}{\beta^2}d\beta = 0, \qquad \alpha d\alpha + \beta d\beta = 0.$$

Two parameters α, β connected by a relation have been introduced; both equations have been differentiated totally with respect to the parameters; and the problem is to eliminate α, β, $d\alpha$, $d\beta$ from the equations. In this case it is simpler to carry both parameters than to introduce the radicals which would be required if only one parameter were used. The elimination of $d\alpha$, $d\beta$ from the last two equations gives $x : y = \alpha^3 : \beta^3$ or $\sqrt[3]{x} : \sqrt[3]{y} = \alpha : \beta$. From this and the first equation,

$$\frac{1}{\alpha} = \frac{1}{x^{\frac{1}{3}}(x^{\frac{2}{3}} + y^{\frac{2}{3}})}, \quad \frac{1}{\beta} = \frac{1}{y^{\frac{1}{3}}(x^{\frac{2}{3}} + y^{\frac{2}{3}})}, \quad \text{and hence} \quad x^{\frac{2}{3}} + y^{\frac{2}{3}} = K^{\frac{2}{3}}.$$

66. Consider two neighboring curves of $F(x, y, \alpha) = 0$. Let (x_0, y_0) be an ordinary point of $\alpha = \alpha_0$ and $(x_0 + dx, y_0 + dy)$ of $\alpha_0 + d\alpha$. Then

$$\begin{aligned} F(x_0 + dx, y_0 + dy, \alpha_0 + d\alpha) - F(x_0, y_0, \alpha_0)& \\ = F'_x dx + F'_y dy + F'_\alpha d\alpha = 0& \qquad (37) \end{aligned}$$

holds except for infinitesimals of higher order. The distance from the point on $\alpha_0 + d\alpha$ to the tangent to α_0 at (x_0, y_0) is

$$\frac{F'_x dx + F'_y dy}{\pm\sqrt{F'^2_x + F'^2_y}} = \frac{\mp F'_\alpha d\alpha}{\sqrt{F'^2_x + F'^2_y}} = dn \qquad (38)$$

except for infinitesimals of higher order. This distance is of the first order with $d\alpha$, and the normal derivative $d\alpha/dn$ of § 48 is finite except when $F'_\alpha = 0$. The distance is of higher order than $d\alpha$, and $d\alpha/dn$ is infinite or $dn/d\alpha$ is zero when $F'_\alpha = 0$. It appears therefore that *the envelope is the locus of points at which the distance between two neighboring curves is of higher order than $d\alpha$*. This is also apparent geometrically from the fact that the distance from a point on a curve to the

tangent to the curve at a neighboring point is of higher order (§ 36). Singular points have been ruled out because (38) becomes indeterminate. In general the locus of singular points is not tangent to the curves of the family and is not an envelope but an extraneous factor; in exceptional cases this locus is an envelope.

If two neighboring curves $F(x, y, \alpha) = 0$, $F(x, y, \alpha + \Delta\alpha) = 0$ intersect, their point of intersection satisfies both of the equations, and hence also the equation

$$\frac{1}{\Delta\alpha}\left[F(x, y, \alpha + \Delta\alpha) - F(x, y, \alpha)\right] = F'_\alpha(x, y, \alpha + \theta\Delta\alpha) = 0.$$

If the limit be taken for $\Delta\alpha \doteq 0$, the limiting position of the intersection satisfies $F'_\alpha = 0$ and hence may lie on the envelope, and will lie on the envelope if the common point of intersection is remote from singular points of the curves $F(x, y, \alpha) = 0$. This idea of an *envelope as the limit of points in which neighboring curves of the family intersect* is valuable. It is sometimes taken as the definition of the envelope. But, unless imaginary points of intersection are considered, it is an inadequate definition; for otherwise $y = (x - \alpha)^3$ would have no envelope according to the definition (whereas $y = 0$ is obviously an envelope) and a curve could not be regarded as the envelope of its osculating circles.

Care must be used in applying the rule for finding an envelope. Otherwise not only may extraneous solutions be mistaken for the envelope, but the envelope may be missed entirely. Consider

$$y - \sin\alpha x = 0 \quad \text{or} \quad \alpha - x^{-1}\sin^{-1} y = 0, \tag{39}$$

where the second form is obtained by solution and contains a multiple valued function. These two families of curves are identical, and it is geometrically clear that they have an envelope, namely $y = \pm 1$. This is precisely what would be found on applying the rule to the first of (39); but if the rule be applied to the second of (39), it is seen that $F'_\alpha = 1$, which does not vanish and hence indicates no envelope. The whole matter should be examined carefully in the light of implicit functions.

Hence let $F(x, y, \alpha) = 0$ be a continuous single valued function of the three variables (x, y, α) and let its derivatives F'_x, F'_y, F'_α exist and be continuous. Consider the behavior of the curves of the family near a point (x_0, y_0) of the curve for $\alpha = \alpha_0$ provided that (x_0, y_0) is an ordinary (nonsingular) point of the curve and that the derivative $F'_\alpha(x_0, y_0, \alpha_0)$ does not vanish. As $F'_\alpha \neq 0$ and either $F'_x \neq 0$ or $F'_y \neq 0$ for (x_0, y_0, α_0), it is possible to surround (x_0, y_0) with a region so small that $F(x, y, \alpha) = 0$ may be solved for $\alpha = f(x, y)$ which will be single valued and differentiable; and the region may further be taken so small that F'_x or F'_y remains different from 0 throughout the region. Then through every point of the region there is one and only one curve $\alpha = f(x, y)$ and the curves have no singular points within the region. In particular no two curves of the family can be tangent to each other within the region.

Furthermore, in such a region there is no envelope. For let any curve which traverses the region be $x = \phi(t)$, $y = \psi(t)$. Then

$$\alpha(t) = f(\phi(t), \psi(t)), \qquad \alpha'(t) = f'_x\phi'(t) + f'_y\psi'(t).$$

Along any curve $\alpha = f(x, y)$ the equation $f'_x dx + f'_y dy = 0$ holds, and if $x = \phi(t)$, $y = \psi(t)$ be tangent to this curve, $dy = dx = \psi' : \phi'$ and $\alpha'(t) = 0$ or $\alpha = \text{const.}$ Hence the only curve which has at each point the direction of the curve of the family through that point is a curve which coincides throughout with some curve of the family and is tangent to no other member of the family. Hence there is no envelope. The result is that an envelope can be present only when $F'_\alpha = 0$ or when $F'_x = F'_y = 0$, and this latter case has been seen to be included in the condition $F'_\alpha = 0$. If $F(x, y, \alpha)$ were not single valued but the branches were separable, the same conclusion would hold. Hence in case $F(x, y, \alpha)$ is not single valued the loci over which two or more values become inseparable must be added to those over which $F'_\alpha = 0$ in order to insure that all the loci which may be envelopes are taken into account.

67. The preceding considerations apply with so little change to other cases of envelopes that the facts will merely be stated without proof. Consider a family of surfaces $F(x, y, z, \alpha, \beta) = 0$ depending on two parameters. The envelope may be defined by the property of tangency as in § 65; and *the conditions for an envelope would be*

$$F(x, y, z, \alpha, \beta) = 0, \qquad F'_\alpha = 0, \qquad F'_\beta = 0. \tag{40}$$

These three equations may be solved to express the envelope as

$$x = \phi(\alpha, \beta), \qquad y = \psi(\alpha, \beta), \qquad z = \chi(\alpha, \beta)$$

parametrically in terms of α, β; or the two parameters may be eliminated and the envelope may be found as $\Phi(x, y, z) = 0$. In any case extraneous loci may be introduced and the results of the work should therefore be tested, which generally may be done at sight.

It is also possible to determine the distance from the tangent plane of one surface to the neighboring surfaces as

$$\frac{F'_x dx + F'_y dy + F'_z dz}{\sqrt{F'^2_x + F'^2_y + F'^2_z}} = \frac{F'_\alpha d\alpha + F'_\beta d\beta}{\sqrt{F'^2_x + F'^2_y + F'^2_z}} = dn, \tag{41}$$

and to define the envelope as the locus of points such that this distance is of higher order than $|d\alpha| + |d\beta|$. The equations (40) would then also follow. This definition would apply only to ordinary points of the surfaces of the family, that is, to points for which not all the derivatives F'_x, F'_y, F'_z vanish. But as the elimination of α, β from (40) would give an equation which included the loci of these singular points, there would be no danger of losing such loci in the rare instances where they, too, happened to be tangent to the surfaces of the family.

The application of implicit functions as in § 66 could also be made in this case and would show that no envelope could exist in regions where no singular points occurred and where either F'_α or F'_β failed to vanish. This work could be based either on the first definition involving tangency directly or on the second definition which involves tangency indirectly in the statements concerning infinitesimals of higher order. It may be added that if $F(x, y, z, \alpha, \beta) = 0$ were not single valued, the surfaces over which two values of the function become inseparable should be added as possible envelopes.

A family of surfaces $F(x, y, z, \alpha) = 0$ depending on a single parameter may have an envelope, and *the envelope is found from*

$$F(x, y, z, \alpha) = 0, \qquad F'_\alpha(x, y, z, \alpha) = 0 \tag{42}$$

by the elimination of the single parameter. The details of the deduction of the rule will be omitted. If two neighboring surfaces intersect, the limiting position of the curve of intersection lies on the envelope and the envelope is the surface generated by this curve as α varies. The surfaces of the family touch the envelope not at a point merely but along these curves. The curves are called *characteristics* of the family. In the case where consecutive surfaces of the family do not intersect in a real curve it is necessary to fall back on the conception of imaginaries or on the definition of an envelope in terms of tangency or infinitesimals; the characteristic curves are still the curves along which the surfaces of the family are in contact with the envelope and along which two consecutive surfaces of the family are distant from each other by an infinitesimal of higher order than $d\alpha$.

A particular case of importance is the envelope of a plane which depends on one parameter. The equations (42) are then

$$Ax + By + Cz + D = 0, \qquad A'x + B'y + C'z + D' = 0, \tag{43}$$

where A, B, C, D are functions of the parameter and differentiation with respect to it is denoted by accents. The case where the plane moves parallel to itself or turns about a line may be excluded as trivial. As the intersection of two planes is a line, the characteristics of the system are straight lines, the envelope is a *ruled surface*, and *a plane tangent to the surface at one point of the lines is tangent to the surface throughout the whole extent of the line.* Cones and cylinders are examples of this sort of surface. Another example is the surface enveloped by the osculating planes of a curve in space; for the osculating plane depends on only one parameter. As the osculating plane (§ 41) may be regarded as passing through three consecutive points of the curve, two consecutive osculating planes may be considered as having two consecutive points of the curve in common and hence the characteristics are

the tangent lines to the curve. Surfaces which are the envelopes of a plane which depends on a single parameter are called *developable surfaces.*

A family of curves dependent on two parameters as

$$F(x, y, z, \alpha, \beta) = 0, \qquad G(x, y, z, \alpha, \beta) = 0 \tag{44}$$

is called a *congruence of curves.* The curves may have an envelope, that is, there may be a surface to which the curves are tangent and which may be regarded as the locus of their points of tangency. The envelope is obtained by eliminating α, β from the equations

$$F = 0, \qquad G = 0, \qquad F'_\alpha G'_\beta - F'_\beta G'_\alpha = 0. \tag{45}$$

To see this, suppose that the third condition is not fulfilled. The equations (44) may then be solved as $\alpha = f(x, y, z)$, $\beta = g(x, y, z)$. Reasoning like that of § 66 now shows that there cannot possibly be an envelope in the region for which the solution is valid. It may therefore be inferred that the only possibilities for an envelope are contained in the equations (45). As various extraneous loci might be introduced in the elimination of α, β from (45) and as the solutions should therefore be tested individually, it is hardly necessary to examine the general question further. The envelope of a congruence of curves is called the *focal surface* of the congruence and the points of contact of the curves with the envelope are called the *focal points* on the curves.

EXERCISES

1. Find the envelopes of these families of curves. In each case test the answer or its individual factors and check the results by a sketch:

(α) $y = 2\alpha x + \alpha^4$, $\qquad(\beta)$ $y^2 = \alpha(x - \alpha)$, $\qquad(\gamma)$ $y = \alpha x + k/\alpha$,

(δ) $\alpha(y + \alpha)^2 = x^3$, $\qquad(\epsilon)$ $y = \alpha(x + \alpha)^2$, $\qquad(\zeta)$ $y^2 = \alpha(x - \alpha)^3$.

2. Find the envelope of the ellipses $x^2/a^2 + y^2/b^2 = 1$ under the condition that (α) the sum of the axes is constant or (β) the area is constant.

3. Find the envelope of the circles whose center is on a given parabola and which pass through the vertex of the parabola.

4. Circles pass through the origin and have their centers on $x^2 - y^2 = c^2$. Find their envelope. *Ans.* A lemniscate.

5. Find the envelopes in these cases:

(α) $x + xy\alpha = \sin^{-1} xy$, $\qquad(\beta)$ $x + \alpha = \operatorname{vers}^{-1} y + \sqrt{2y - y^2}$,

(γ) $y + \alpha = \sqrt{1 - 1/x}$.

6. Find the envelopes in these cases:

(α) $\alpha x + \beta y + \alpha\beta z = 1$, $\qquad(\beta)$ $\dfrac{x}{\alpha} + \dfrac{y}{\beta} + \dfrac{z}{1 - \alpha - \beta} = 1$,

(γ) $\dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} + \dfrac{z^2}{\gamma^2} = 1$ with $\alpha\beta\gamma = k^3$.

7. Find the envelopes in Ex. 6 (α), (β) if $\alpha = \beta$ or if $\alpha = -\beta$.

8. Prove that the envelope of $F(x, y, z, \alpha) = 0$ is tangent to the surface along the whole characteristic by showing that the normal to $F(x, y, z, \alpha) = 0$ and to the eliminant of $F = 0$, $F'_\alpha = 0$ are the same, namely

$$F'_x : F'_y : F'_z \quad \text{and} \quad F'_x + F'_\alpha \frac{\partial \alpha}{\partial x} : F'_y + F'_\alpha \frac{\partial \alpha}{\partial y} : F'_z + F'_\alpha \frac{\partial \alpha}{\partial z},$$

where $\alpha(x, y, z)$ is the function obtained by solving $F'_\alpha = 0$. Consider the problem also from the point of view of infinitesimals and the normal derivative.

9. If there is a curve $x = \phi(\alpha)$, $y = \psi(\alpha)$, $z = \chi(\alpha)$ tangent to the curves of the family defined by $F(x, y, z, \alpha) = 0$, $G(x, y, z, \alpha) = 0$ in space, then that curve is called the envelope of the family. Show, by the same reasoning as in § 65 for the case of the plane, that the four conditions $F = 0$, $G = 0$, $F'_\alpha = 0$, $G'_\alpha = 0$ must be satisfied for an envelope; and hence infer that ordinarily a family of curves in space dependent on a single parameter has no envelope.

10. Show that the family $F(x, y, z, \alpha) = 0$, $F'_\alpha(x, y, z, \alpha) = 0$ of curves which are the characteristics of a family of surfaces has in general an envelope given by the three equations $F = 0$, $F'_\alpha = 0$, $F''_{\alpha\alpha} = 0$.

11. Derive the condition (45) for the envelope of a two-parametered family of curves from the idea of tangency, as in the case of one parameter.

12. Find the envelope of the normals to a plane curve $y = f(x)$ and show that the envelope is the locus of the center of curvature.

13. The locus of Ex. 12 is called the *evolute* of the curve $y = f(x)$. In these cases find the evolute as an envelope:

(α) $y = x^2$, (β) $x = a \sin t$, $y = b \cos t$, (γ) $2xy = a^2$,
(δ) $y^2 = 2mx$, (ϵ) $x = a(\theta - \sin\theta)$, $y = a(1 - \cos\theta)$, (ζ) $y = \cosh x$.

14. Given a surface $z = f(x, y)$. Construct the family of normal lines and find their envelope.

15. If rays of light issuing from a point in a plane are reflected from a curve in the plane, the angle of reflection being equal to the angle of incidence, the envelope of the reflected rays is called the *caustic* of the curve with respect to the point. Show that the caustic of a circle with respect to a point on its circumference is a cardioid.

16. The curve which is the envelope of the characteristic lines, that is, of the rulings, on the developable surface (43) is called the *cuspidal edge* of the surface. Show that the equations of this curve may be found parametrically in terms of the parameter of (43) by solving simultaneously

$$Ax + By + Cz + D = 0,\ A'x + B'y + C'z + D' = 0,\ A''x + B''y + C''z + D'' = 0$$

for x, y, z. Consider the exceptional cases of cones and cylinders.

17. The term "developable" signifies that a *developable surface may be developed or mapped on a plane in such a way that lengths of arcs on the surface become equal lengths in the plane*, that is, the map may be made without distortion of size or shape. In the case of cones or cylinders this map may be made by slitting the cone or cylinder along an element and rolling it out upon a plane. What is the analytic statement in this case? In the case of any developable surface with a cuspidal edge, the developable surface being the locus of all tangents to the cuspidal edge,

the length of arc upon the surface may be written as $d\sigma^2 = (dt + ds)^2 + t^2ds^2/R^2$, where s denotes arc measured along the cuspidal edge and t denotes distance along the tangent line. This form of $d\sigma^2$ may be obtained geometrically by infinitesimal analysis or analytically from the equations

$$x = f(s) + tf'(s), \quad y = g(s) + tg'(s), \quad z = h(s) + th'(s)$$

of the developable surface of which $x = f(s)$, $y = g(s)$, $z = h(s)$ is the cuspidal edge. It is thus seen that $d\sigma^2$ is the same at corresponding points of all developable surfaces for which the radius of curvature R of the cuspidal edge is the same function of s without regard to the torsion; in particular the torsion may be zero and the developable may reduce to a plane.

18. Let the line $x = az + b$, $y = cz + d$ depend on one parameter so as to generate a ruled surface. By identifying this form of the line with (43) obtain by substitution the conditions

$$\begin{matrix} Aa + Bc + C = 0, & A'a + B'c + C' = 0 \\ Ab + Bd + D = 0, & A'b + B'd + D' = 0 \end{matrix} \quad \text{or} \quad \begin{matrix} Aa' + Bc' = 0 \\ Ab' + Bd' = 0 \end{matrix} \quad \text{or} \quad \begin{vmatrix} a' & c' \\ b' & d' \end{vmatrix} = 0$$

as the condition that the line generates a developable surface.

68. More differential geometry. The representation

$$F(x, y, z) = 0, \quad \text{or} \quad z = f(x, y) \tag{46}$$

or $$x = \phi(u, v), \qquad y = \psi(u, v), \qquad z = \chi(u, v)$$

of a surface may be taken in the unsolved, the solved, or the parametric form. The parametric form is equivalent to the solved form provided u, v be taken as x, y. The notation

$$p = \frac{\partial z}{\partial x}, \qquad q = \frac{\partial z}{\partial y}, \qquad r = \frac{\partial^2 z}{\partial x^2}, \qquad s = \frac{\partial^2 z}{\partial x \partial y}, \qquad t = \frac{\partial^2 z}{\partial y^2}$$

is adopted for the derivatives of z with respect to x and y. The application of Taylor's Formula to the solved form gives

$$\Delta z = ph + qk + \tfrac{1}{2}(rh^2 + 2shk + tk^2) + \cdots \tag{47}$$

with $h = \Delta x$, $k = \Delta y$. The linear terms $ph + qk$ constitute the differential dz and represent that part of the increment of z which would be obtained by replacing the surface by its tangent plane. Apart from infinitesimals of the third order, the distance from the tangent plane up or down to the surface along a parallel to the z-axis is given by the quadratic terms $\frac{1}{2}(rh^2 + 2shk + tk^2)$.

Hence if the quadratic terms at any point are a positive definite form (§ 55), the surface lies above its tangent plane and is concave up; but if the form is negative definite, the surface lies below its tangent plane and is concave down or convex up. If the form is indefinite but not singular, the surface lies partly above and partly below its tangent plane and may be called concavo-convex, that is, it is saddle-shaped. If the form is singular nothing can be definitely stated. These statements

are merely generalizations of those of § 55 made for the case where the tangent plane is parallel to the xy-plane. It will be assumed in the further work of these articles that at least one of the derivatives r, s, t is not 0.

To examine more closely the behavior of a surface in the vicinity of a particular point upon it, let the xy-plane be taken in coincidence with the tangent plane at the point and let the point be taken as origin. Then Maclaurin's Formula is available.

$$\begin{aligned} z &= \tfrac{1}{2}(rx^2 + 2sxy + ty^2) + \text{terms of higher order} \\ &= \tfrac{1}{2}\rho^2(r\cos^2\theta + 2s\sin\theta\cos\theta + t\sin^2\theta) + \text{higher terms,} \end{aligned} \tag{48}$$

where (ρ, θ) are polar coördinates in the xy-plane. Then

$$\frac{1}{R} = r\cos^2\theta + 2s\sin\theta\cos\theta + t\sin^2\theta = \frac{d^2z}{d\rho^2} \div \left[1 + \left(\frac{dz}{d\rho}\right)^2\right]^{\frac{3}{2}} \tag{49}$$

is the curvature of a normal section of the surface. The sum of the curvatures in two normal sections which are in perpendicular planes may be obtained by giving θ the values θ and $\theta + \frac{1}{2}\pi$. This sum reduces to $r + t$ and is therefore independent of θ.

As the sum of the curvatures in two perpendicular normal planes is constant, the maximum and minimum values of the curvature will be found in perpendicular planes. These values of the curvature are called the *principal values* and their reciprocals are the *principal radii of curvature* and the sections in which they lie are the *principal sections*. If $s = 0$, the principal sections are $\theta = 0$ and $\theta = \frac{1}{2}\pi$; and conversely if the axes of x and y had been chosen in the tangent plane so as to be tangent to the principal sections, the derivative s would have vanished. The equation of the surface would then have taken the simple form

$$z = \tfrac{1}{2}(rx^2 + ty^2) + \text{higher terms.} \tag{50}$$

The principal curvatures would be merely r and t, and the curvature in any normal section would have had the form

$$\frac{1}{R} = \frac{\cos^2\theta}{R_1} + \frac{\sin^2\theta}{R_2} = r\cos^2\theta + t\sin^2\theta.$$

If the two principal curvatures have opposite signs, that is, if the signs of r and t in (50) are opposite, the surface is saddle-shaped. There are then two directions for which the curvature of a normal section vanishes, namely the directions of the lines

$$\theta = \pm\tan^{-1}\sqrt{-R_2/R_1} \quad \text{or} \quad \sqrt{|r|}\,x = \pm\sqrt{|t|}\,y.$$

These are called the *asymptotic directions*. Along these directions the surface departs from its tangent plane by infinitesimals of the third

order, or higher order. If a curve is drawn on a surface so that at each point of the curve the tangent to the curve is along one of the asymptotic directions, the curve is called an *asymptotic curve or line* of the surface. As the surface departs from its tangent plane by infinitesimals of higher order than the second along an asymptotic line, the tangent plane to a surface at any point of an asymptotic line must be the osculating plane of the asymptotic line.

The character of a point upon a surface is indicated by the *Dupin indicatrix* of the point. The indicatrix is the conic

$$\frac{x^2}{R_1} + \frac{y^2}{R_2} = 1, \qquad \text{cf. } z = \tfrac{1}{2}(rx^2 + ty^2), \tag{51}$$

which has the principal directions as the directions of its axes and the square roots of the absolute values of the principal radii of curvature as the magnitudes of its axes. The conic may be regarded as similar to the conic in which a plane infinitely near the tangent plane cuts the surface when infinitesimals of order higher than the second are neglected. In case the surface is concavo-convex the indicatrix is a hyperbola and should be considered as either or both of the two conjugate hyperbolas that would arise from giving z positive or negative values in (51). The point on the surface is called elliptic, hyperbolic, or parabolic according as the indicatrix is an ellipse, a hyperbola, or a pair of lines, as happens when one of the principal curvatures vanishes. These classes of points correspond to the distinctions definite, indefinite, and singular applied to the quadratic form $rh^2 + 2shk + tk^2$.

Two further results are noteworthy. Any curve drawn on the surface differs from the section of its osculating plane with the surface by infinitesimals of higher order than the second. For as the osculating plane passes through three consecutive points of the curve, its intersection with the surface passes through the same three consecutive points and the two curves have contact of the second order. It follows that the radius of curvature of any curve on the surface is identical with that of the curve in which its osculating plane cuts the surface. The other result is *Meusnier's Theorem:* The radius of curvature of an oblique section of the surface at any point is the projection upon the plane of that section of the radius of curvature of the normal section which passes through the same tangent line. In other words, if the radius of curvature of a normal section is known, that of the oblique sections through the same tangent line may be obtained by multiplying it by the cosine of the angle between the plane normal to the surface and the plane of the oblique section.

The proof of Meusnier's Theorem may be given by reference to (48). Let the x-axis in the tangent plane be taken along the intersection with the oblique plane. Neglect infinitesimals of higher order than the second. Then

$$y = \phi(x) = \tfrac{1}{2} a x^2, \qquad z = \tfrac{1}{2}(rx^2 + 2sxy + ty^2) = \tfrac{1}{2} rx^2 \tag{48'}$$

will be the equations of the curve. The plane of the section is $az - ry = 0$, as may be seen by inspection. The radius of curvature of the curve in this plane may be found at once. For if u denote distance in the plane and perpendicular to the x-axis and if ν be the angle between the normal plane and the oblique plane $az - ry = 0$,

$$u = z \sec\nu = y \csc\nu = \tfrac{1}{2} r \sec\nu \cdot x^2 = \tfrac{1}{2} a \csc\nu \cdot x^2.$$

The form $u = \frac{1}{2} r \sec\nu \cdot x^2$ gives the curvature as $r \sec\nu$. But the curvature in the normal section is r by (48′). As the curvature in the oblique section is $\sec\nu$ times that in the normal section, the radius of curvature in the oblique section is $\cos\nu$ times that of the normal section. Meusnier's Theorem is thus proved.

69. These investigations with a special choice of axes give geometric properties of the surface, but do not express those properties in a convenient analytic form; for if a surface $z = f(x, y)$ is given, the transformation to the special axes is difficult. The idea of the indicatrix or its similar conic as the section of the surface by a plane near the tangent plane and parallel to it will, however, determine the general conditions readily. If in the expansion

$$\Delta z - dz = \tfrac{1}{2}(rh^2 + 2shk + tk^2) = \text{const.} \tag{52}$$

the quadratic terms be set equal to a constant, the conic obtained is the projection of the indicatrix on the xy-plane, or if (52) be regarded as a cylinder upon the xy-plane, the indicatrix (or similar conic) is the intersection of the cylinder with the tangent plane. As the character of the conic is unchanged by the projection, *the point on the surface is elliptic if $s^2 < rt$, hyperbolic if $s^2 > rt$, and parabolic if $s^2 = rt$.* Moreover if the indicatrix is hyperbolic, its asymptotes must project into the asymptotes of the conic (52), and hence if dx and dy replace h and k, the equation

$$rdx^2 + 2sdxdy + tdy^2 = 0 \tag{53}$$

may be regarded as *the differential equation of the projection of the asymptotic lines on the xy-plane.* If r, s, t be expressed as functions $f''_{xx}, f''_{xy}, f''_{yy}$ of (x, y) and (53) be factored, the integration of the two equations $M(x, y)\,dx + N(x, y)\,dy$ thus found will give the finite equations of the projections of the asymptotic lines and, taken with the equation $z = f(x, y)$, will give the curves on the surface.

To find the lines of curvature is not quite so simple; for it is necessary to determine the directions which are the projections of the axes of the indicatrix, and these are not the axes of the projected conic. Any radius of the indicatrix may be regarded as the intersection of the tangent plane and a plane perpendicular to the xy-plane through the radius of the projected conic. Hence

$$z - z_0 = p(x - x_0) + q(y - y_0), \qquad (x - x_0)k = (y - y_0)h$$

are the two planes which intersect in the radius that projects along the direction determined by h, k. The direction cosines

$$\frac{h : k : ph + qk}{\sqrt{h^2 + k^2 + (ph + qk)^2}} \quad \text{and} \quad h : k : 0 \tag{54}$$

are therefore those of the radius in the indicatrix and of its projection and they determine the cosine of the angle ϕ between the radius and its projection. The square of the radius in (52) is

$$h^2 + k^2, \quad \text{and} \quad (h^2 + k^2)\sec^2\phi = h^2 + k^2 + (ph + qk)^2$$

is therefore the square of the corresponding radius in the indicatrix. To determine the axes of the indicatrix, this radius is to be made a maximum or minimum subject to (52). With a multiplier λ,

$$h + ph + qk + \lambda(rh + sk) = 0, \qquad k + ph + qk + \lambda(sh + tk) = 0$$

are the conditions required, and the elimination of λ gives

$$h^2[s(1 + p^2) - pqr] + hk[t(1 + p^2) - r(1 + q^2)] - k^2[t(1 + q^2) - pqt] = 0$$

as the equation that determines the projection of the axes. Or

$$\frac{(1 + p^2)dx + pqdy}{rdx + sdy} = \frac{pqdx + (1 + q^2)dy}{sdx + tdy} \tag{55}$$

is the differential equation of the projected lines of curvature.

In addition to the asymptotic lines and lines of curvature the *geodesic or shortest lines* on the surface are important. These, however, are better left for the methods of the calculus of variations (§ 159). The attention may therefore be turned to finding the value of the radius of curvature in any normal section of the surface.

A reference to (48) and (49) shows that the curvature is

$$\frac{1}{R} = \frac{2z}{\rho^2} = \frac{rh^2 + 2shk + tk^2}{\rho^2} = \frac{rh^2 + 2shk + tk^2}{h^2 + k^2}$$

in the special case. But in the general case the normal distance to the surface is $(\Delta z - dz)\cos\gamma$, with $\sec\gamma = \sqrt{1 + p^2 + q^2}$, instead of the $2z$ of the special case, and the radius ρ^2 of the special case becomes $\rho^2\sec^2\phi = h^2 + k^2 + (ph + qk)^2$ in the tangent plane. Hence

$$\frac{1}{R} = \frac{2(\Delta z - dz)\cos\gamma}{h^2 + k^2 + (ph + qk)^2} = \frac{rl^2 + 2slm + tm^2}{\sqrt{1 + p^2 + q^2}}, \tag{56}$$

where the direction cosines l, m of a radius in the tangent plane have been introduced from (54), is the general expression for the curvature of a normal section. The form

$$\frac{1}{R} = \frac{rh^2 + 2shk + tk^2}{h^2 + k^2 + (ph + qk)^2}\,\frac{1}{\sqrt{1 + p^2 + q^2}}, \tag{56'}$$

where the direction h, k of the projected radius remains, is frequently more convenient than (56) which contains the direction cosines l, m of the original direction in the tangent plane. Meusnier's Theorem may now be written in the form

$$\frac{\cos\nu}{R} = \frac{rl^2 + 2slm + tm^2}{\sqrt{1 + p^2 + q^2}}, \tag{57}$$

where ν is the angle between an oblique section and the tangent plane and where l, m are the direction cosines of the intersection of the planes.

The work here given has depended for its relative simplicity of statement upon the assumption of the surface (46) in solved form. It is merely a problem in implicit partial differentiation to pass from p, q, r, s, t to their equivalents in terms of F'_x, F'_y, F'_z or the derivatives of ϕ, ψ, χ by α, β.

EXERCISES

1. In (49) show $\dfrac{1}{R} = \dfrac{r+t}{2} + \dfrac{r-t}{2}\cos 2\theta + s\sin 2\theta$ and find the directions of maximum and minimum R. If R_1 and R_2 are the maximum and minimum values of R, show

$$\frac{1}{R_1} + \frac{1}{R_2} = r + t \quad \text{and} \quad \frac{1}{R_1}\cdot\frac{1}{R_2} = rt - s^2.$$

Half of the sum of the curvatures is called the *mean curvature*; the product of the curvatures is called the *total curvature*.

2. Find the mean curvature, the total curvature, and therefrom (by constructing and solving a quadratic equation) the principal radii of curvature at the origin:

$$(\alpha)\ z = xy, \qquad (\beta)\ z = x^2 + xy + y^2, \qquad (\gamma)\ z = x(x+y).$$

3. In the surfaces (α) $z = xy$ and (β) $z = 2x^2 + y^2$ find at (0, 0) the radius of curvature in the sections made by the planes

$$(\alpha)\ x + y = 0, \qquad (\beta)\ x + y + z = 0, \qquad (\gamma)\ x + y + 2z = 0.$$
$$(\delta)\ x - 2y = 0, \qquad (\epsilon)\ x - 2y + z = 0, \qquad (\zeta)\ x + 2y + \tfrac{1}{2}z = 0.$$

The oblique sections are to be treated by applying Meusnier's Theorem.

4. Find the asymptotic directions at (0, 0) in Exs. 2 and 3.

5. Show that *a developable surface is everywhere parabolic*, that is, that $rt - s^2 = 0$ at every point; and conversely. To do this consider the surface as the envelope of its tangent plane $z - p_0x - q_0y = z_0 - p_0x_0 - q_0y_0$, where p_0, q_0, x_0, y_0, z_0 are functions of a single parameter α. Hence show

$$J\begin{pmatrix} p_0,\ q_0 \\ x_0,\ y_0 \end{pmatrix} = 0 = (rt - s^2)_0 \quad \text{and} \quad J\begin{pmatrix} p_0,\ z_0 - p_0x_0 - q_0y_0 \\ x_0,\ y_0 \end{pmatrix} = y_0(s^2 - rt)_0.$$

The first result proves the statement; the second, its converse.

6. Find the differential equations of the asymptotic lines and lines of curvature on these surfaces:

$$(\alpha)\ z = xy, \qquad (\beta)\ z = \tan^{-1}(y/x), \qquad (\gamma)\ z^2 + y^2 = \cosh x, \qquad (\delta)\ xyz = 1.$$

7. Show that the mean curvature and total curvature are

$$\frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{r(1+q^2) + t(1+p^2) - 2pqs}{2(1+p^2+q^2)^{\frac{3}{2}}}, \qquad \frac{1}{R_1R_2} = \frac{rt - s^2}{(1+p^2+q^2)^2}.$$

8. Find the principal radii of curvature at (1, 1) in Ex. 6.

9. An umbilic is a point of a surface at which the principal radii of curvature (and hence all radii of curvature for normal sections) are equal. Show that the conditions are $\dfrac{r}{1+p^2} = \dfrac{s}{pq} = \dfrac{t}{1+q^2}$ for an umbilic, and determine the umbilics of the ellipsoid with semiaxes a, b, c.

CHAPTER VI

COMPLEX NUMBERS AND VECTORS

70. Operators and operations. If an entity u is changed into an entity v by some law, the change may be regarded as an *operation* performed upon u, the *operand*, to convert it into v; and if f be introduced as the symbol of the operation, the result may be written as $v = fu$. For brevity the symbol f is often called an *operator*. Various sorts of operand, operator, and result are familiar. Thus if u is a positive number n, the application of the operator $\sqrt{\ }$ gives the square root; if u represents a range of values of a variable x, the expression $f(x)$ or fx denotes a function of x; if u be a function of x, the operation of differentiation may be symbolized by D and the result Du is the derivative; the symbol of definite integration $\int_a^b (*)\,d*$ converts a function $u(x)$ into a number; and so on in great variety.

The reason for making a short study of operators is that a considerable number of the concepts and rules of arithmetic and algebra may be so defined for operators themselves as to lead to a *calculus of operations* which is of frequent use in mathematics; the single application to the integration of certain differential equations (§ 95) is in itself highly valuable. The fundamental concept is that of a *product: If u is operated upon by f to give $fu = v$ and if v is operated upon by g to give $gv = w$, so that*

$$fu = v, \qquad gv = gfu = w, \qquad gfu = w, \tag{1}$$

then the operation indicated as gf which converts u directly into w is called the product of f by g. If the functional symbols sin and log be regarded as operators, the symbol log sin could be regarded as the product. The transformations of turning the xy-plane over on the x-axis, so that $x' = x$, $y' = -y$, and over the y-axis, so that $x' = -x$, $y' = y$, may be regarded as operations; the combination of these operations gives the transformation $x' = -x$, $y' = -y$, which is equivalent to rotating the plane through 180° about the origin.

The products of arithmetic and algebra satisfy the *commutative law* $gf = fg$, that is, the products of g by f and of f by g are equal. This is not true of operators in general, as may be seen from the fact that

$\log \sin x$ and $\sin \log x$ are different. Whenever the order of the factors is immaterial, as in the case of the transformations just considered, the operators are said to be *commutative*. Another law of arithmetic and algebra is that when there are three or more factors in a product, the factors may be grouped at pleasure without altering the result, that is,

$$h(gf) = (hg)f = hgf. \tag{2}$$

This is known as the *associative law* and operators which obey it are called *associative*. Only associative operators are considered in the work here given.

For the repetition of an operator several times

$$ff = f^2, \quad fff = f^3, \quad f^m f^n = f^{m+n}, \tag{3}$$

the usual notation of powers is used. *The law of indices clearly holds;* for f^{m+n} means that f is applied $m + n$ times successively, whereas $f^m f^n$ means that it is applied n times and then m times more. Not applying the operator f at all would naturally be denoted by f^0, so that $f^0 u = u$ and the operator f^0 would be equivalent to multiplication by 1; the notation $f^0 = 1$ is adopted.

If for a given operation f there can be found an operation g such that the product $fg = f^0 = 1$ is equivalent to no operation, then g is called the *inverse* of f and notations such as

$$fg = 1, \quad g = f^{-1} = \frac{1}{f}, \quad ff^{-1} = f\frac{1}{f} = 1 \tag{4}$$

are regularly borrowed from arithmetic and algebra. Thus the inverse of the square is the square root, the inverse of sin is $\sin^{-1}$, the inverse of the logarithm is the exponential, the inverse of D is $\int$. Some operations have no inverse; multiplication by 0 is a case, and so is the square when applied to a negative number if only real numbers are considered. Other operations have more than one inverse; integration, the inverse of D, involves an arbitrary additive constant, and the inverse sine is a multiple valued function. It is therefore not always true that $f^{-1}f = 1$, but it is customary to mean by f^{-1} that particular inverse of f for which $f^{-1}f = ff^{-1} = 1$. Higher negative powers are defined by the equation $f^{-n} = (f^{-1})^n$, and it readily follows that $f^n f^{-n} = 1$, as may be seen by the example

$$f^3 f^{-3} = ff(f \cdot f^{-1})f^{-1}f^{-1} = f(f \cdot f^{-1})f^{-1} = ff^{-1} = 1.$$

The law of indices $f^m f^n = f^{m+n}$ also holds for negative indices, except in so far as $f^{-1}f$ may not be equal to 1 and may be required in the reduction of $f^m f^n$ to f^{m+n}.

If u, v, and $u + v$ are operands for the operator f and if

$$f(u + v) = fu + fv, \tag{5}$$

so that the operator applied to the sum gives the same result as the sum of the results of operating on each operand, then the operator f is called *linear* or *distributive.* If f denotes a function such that $f(x + y) = f(x) + f(y)$, it has been seen (Ex. 9, p. 45) that f must be equivalent to multiplication by a constant and $fx = Cx$. For a less specialized interpretation this is not so; for

$$D(u + v) = Du + Dv \quad \text{and} \quad \int (u + v) = \int u + \int v$$

are two of the fundamental formulas of calculus and show operators which are distributive and not equivalent to multiplication by a constant. Nevertheless it does follow by the same reasoning as used before (Ex. 9, p. 45), that $fnu = nfu$ if f is distributive and if n is a rational number.

Some operators have also the property of addition. Suppose that u is an operand and f, g are operators such that fu and gu are things that may be added together as $fu + gu$, then the *sum* of the operators, $f + g$, is defined by the equation $(f + g)u = fu + gu$. If furthermore the operators f, g, h are distributive, then

$$h(f + g) = hf + hg \quad \text{and} \quad (f + g)h = fh + gh, \tag{6}$$

and the multiplication of the operators becomes itself distributive. To prove this fact, it is merely necessary to consider that

$$h[(f + g)u] = h(fu + gu) = hfu + hgu$$

and
$$(f + g)(hu) = fhu + ghu.$$

Operators which are associative, commutative, distributive, and which admit addition may be treated algebraically, in so far as polynomials are concerned, by the ordinary algorisms of algebra : for it is by means of the associative, commutative, and distributive laws, and the law of indices that ordinary algebraic polynomials are rearranged, multiplied out, and factored. Now the operations of multiplication by constants and of differentiation or partial differentiation as applied to a function of one or more variables x, y, z, $\cdots$ do satisfy these laws. For instance

$$c(Du) = D(cu), \quad D_x D_y u = D_y D_x u, \quad (D_x + D_y) D_z u = D_x D_z u + D_y D_z u. \tag{7}$$

Hence, for example, if y be a function of x, the expression

$$D^n y + a_1 D^{n-1} y + \cdots + a_{n-1} D y + a_n y,$$

where the coefficients a are constants, may be written as

$$(D^n + a_1 D^{n-1} + \cdots + a_{n-1} D + a_n) y \tag{8}$$

and may then be factored into the form

$$[(D-\alpha_1)(D-\alpha_2)\cdots(D-\alpha_{n-1})(D-\alpha_n)]y, \tag{8'}$$

where $\alpha_1, \alpha_2, \cdots, \alpha_n$ are the roots of the algebraic polynomial

$$x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0.$$

EXERCISES

1. Show that $(fgh)^{-1} = h^{-1}g^{-1}f^{-1}$, that is, that the reciprocal of a product of operations is the product of the reciprocals in inverse order.

2. By definition the operator gfg^{-1} is called the transform of f by g. Show that (α) the transform of a product is the product of the transforms of the factors taken in the same order, and (β) the transform of the inverse is the inverse of the transform.

3. If $s \neq 1$ but $s^2 = 1$, the operator s is by definition said to be *involutory*. Show that (α) an involutory operator is equal to its own inverse; and conversely (β) if an operator and its inverse are equal, the operator is involutory; and (γ) if the product of two involutory operators is commutative, the product is itself involutory; and conversely (δ) if the product of two involutory operators is involutory, the operators are commutative.

4. If f and g are both distributive, so are the products fg and gf.

5. If f is distributive and n rational, show $fnu = nfu$.

6. Expand the following operators first by ordinary formal multiplication and second by applying the operators successively as indicated, and show the results are identical by translating both into familiar forms.

$$(\alpha)\ (D-1)(D-2)y. \quad \textit{Ans.}\ \frac{d^2y}{dx^2} - 3\frac{dy}{dx} + 2y.$$

$$(\beta)\ (D-1)D(D+1)y. \qquad (\gamma)\ D(D-2)(D+1)(D+3)y.$$

7. Show that $(D-a)\left[e^{ax}\int e^{-ax}X\,dx\right] = X$, where X is a function of x, and hence infer that $e^{ax}\int e^{-ax}(*)\,dx$ is the inverse of the operator $(D-a)(*)$.

8. Show that $D(e^{ax}y) = e^{ax}(D+a)y$ and hence generalize to show that if $P(D)$ denote any polynomial in D with constant coefficients, then

$$P(D)\cdot e^{ax}y = e^{ax}P(D+a)y.$$

Apply this to the following and check the results.

$$(\alpha)\ (D^2-3D+2)e^{2x}y = e^{2x}(D^2+D)y = e^{2x}\left(\frac{d^2y}{dx^2} + \frac{dy}{dx}\right).$$

$$(\beta)\ (D^2-3D-2)e^{x}y. \qquad (\gamma)\ (D^3-3D+2)e^{x}y.$$

9. If y is a function of x and $x = e^t$ show that

$$D_xy = e^{-t}D_ty,\ D_x^2y = e^{-2t}D_t(D_t-1)y, \cdots,\ D_x^py = e^{-pt}D_t(D_t-1)\cdots(D_t-p+1)y.$$

10. Is the expression $(hD_x + kD_y)^n$, which occurs in Taylor's Formula (§ 54), the nth power of the operator $hD_x + kD_y$ or is it merely a conventional symbol? The same question relative to $(xD_x + yD_y)^k$ occurring in Euler's Formula (§ 53)?

71. Complex numbers. In the formal solution of the equation $ax^2 + bx + c = 0$, where $b^2 < 4ac$, numbers of the form $m + n\sqrt{-1}$, where m and n are real, arise. Such numbers are called *complex* or *imaginary*: the part m is called the *real part* and $n\sqrt{-1}$ the *pure imaginary part* of the number. It is customary to write $\sqrt{-1} = i$ and to treat i as a literal quantity subject to the relation $i^2 = -1$. The definitions for the *equality*, *addition*, and *multiplication* of complex numbers are

$$\begin{aligned} a + bi = c + di \quad &\text{if and only if} \quad a = c,\ b = d, \\ [a + bi] + [c + di] &= (a + c) + (b + d)\,i, \\ [a + bi]\,[c + di] &= (ac - bd) + (ad + bc)\,i. \end{aligned} \tag{9}$$

It readily follows that *the commutative, associative, and distributive laws hold in the domain of complex numbers*, namely,

$$\begin{aligned} \alpha + \beta &= \beta + \alpha, & (\alpha + \beta) + \gamma &= \alpha + (\beta + \gamma), \\ \alpha\beta &= \beta\alpha, & (\alpha\beta)\,\gamma &= \alpha\,(\beta\gamma), \\ \alpha\,(\beta + \gamma) &= \alpha\beta + \alpha\gamma, & (\alpha + \beta)\,\gamma &= \alpha\gamma + \beta\gamma, \end{aligned} \tag{10}$$

where Greek letters have been used to denote complex numbers.

Division is accomplished by the method of rationalization.

$$\frac{a + bi}{c + di} = \frac{a + bi}{c + di}\cdot\frac{c - di}{c - di} = \frac{(ac + bd) + (bc - ad)\,i}{c^2 + d^2}. \tag{11}$$

This is always possible except when $c^2 + d^2 = 0$, that is, when both c and d are 0. A complex number is defined as 0 when and only when its real and pure imaginary parts are both zero. With this definition 0 has the ordinary properties that $\alpha + 0 = \alpha$ and $\alpha 0 = 0$ and that $\alpha/0$ is impossible. Furthermore *if a product $\alpha\beta$ vanishes, either α or β vanishes.* For suppose

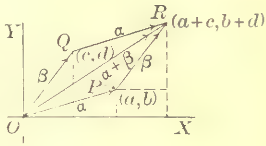
$$[a + bi]\,[c + di] = (ac - bd) + (ad + bc)\,i = 0.$$

Then
$$ac - bd = 0 \quad \text{and} \quad ad + bc = 0, \tag{12}$$

from which it follows that either $a = b = 0$ or $c = d = 0$. From the fact that a product cannot vanish unless one of its factors vanishes follow the ordinary laws of cancellation. In brief, *all the elementary laws of real algebra hold also for the algebra of complex numbers.*

By assuming a set of Cartesian coördinates in the xy-plane and associating the number $a + bi$ to the point (a, b), a *graphical representation* is obtained which is the counterpart of the number scale for real numbers. The point (a, b) alone or the directed line from the origin to the point (a, b) may be considered as representing the number $a + bi$. If OP and OQ are two directed lines representing the two numbers $a + bi$ and $c + di$, a reference to the figure shows that the line which

represents the sum of the numbers is OR, the diagonal of the parallelogram of which OP and OQ are sides. Thus *the geometric law for adding complex numbers is the same as the law for compounding forces and is known as the parallelogram law.* A segment AB of a line possesses magnitude, the length AB, and direction, the direction of the line AB from A to B. *A quantity which has magnitude and direction is called a vector; and the parallelogram law is called the law of vector addition. Complex numbers* may therefore be regarded as *vectors.*



From the figure it also appears that OQ and PR have the same magnitude and direction, so that as vectors they are equal although they start from different points. As $OP + PR$ will be regarded as equal to $OP + OQ$, the definition of addition may be given as the triangle law instead of as the parallelogram law; namely, from the terminal end P of the first vector lay off the second vector PR and close the triangle by joining the initial end O of the first vector to the terminal end R of the second. The *absolute value* of a complex number $a + bi$ is the magnitude of its vector OP and is equal to $\sqrt{a^2 + b^2}$, the square root of the sum of the squares of its real part and of the coefficient of its pure imaginary part. The absolute value is denoted by $|a + bi|$ as in the case of reals. If α and β are two complex numbers, the rule $|\alpha| + |\beta| \geqq |\alpha + \beta|$ is a consequence of the fact that one side of a triangle is less than the sum of the other two. If the absolute value is given and the initial end of the vector is fixed, the terminal end is thereby constrained to lie upon a circle concentric with the initial end.

72. When the complex numbers are laid off from the origin, polar coördinates may be used in place of Cartesian. Then

$$r = \sqrt{a^2 + b^2}, \quad \phi = \tan^{-1} b/a^*, \quad a = r\cos\phi, \quad b = r\sin\phi \tag{13}$$

and $$a + ib = r(\cos\phi + i\sin\phi).$$

The absolute value r is often called the *modulus* or *magnitude* of the complex number; the angle ϕ is called the *angle* or *argument* of the number and suffers a certain indetermination in that $2n\pi$, where n is a positive or negative integer, may be added to ϕ without affecting the number. This polar representation is particularly useful in discussing products and quotients. For if

$$\alpha = r_1(\cos\phi_1 + i\sin\phi_1), \quad \beta = r_2(\cos\phi_2 + i\sin\phi_2), \tag{14}$$

then $$\alpha\beta = r_1 r_2[\cos(\phi_1 + \phi_2) + i\sin(\phi_1 + \phi_2)].$$

* As both $\cos\phi$ and $\sin\phi$ are known, the quadrant of this angle is determined.

as may be seen by multiplication according to the rule. Hence the *magnitude of a product is the product of the magnitudes of the factors, and the angle of a product is the sum of the angles of the factors;* the general rule being proved by induction.

The interpretation of *multiplication by a complex number as an operation* is illuminating. Let β be the multiplicand and α the multiplier. As the product $\alpha\beta$ has a magnitude equal to the product of the magnitudes and an angle equal to the sum of the angles, the factor α used as a multiplier may be interpreted as effecting the rotation of β through the angle of α and the stretching of β in the ratio $|\alpha|:1$. From the geometric viewpoint, therefore, *multiplication by a complex number is an operation of rotation and stretching in the plane.* In the case of $\alpha = \cos\phi + i\sin\phi$ with $r = 1$, the operation is only of rotation and hence the factor $\cos\phi + i\sin\phi$ is often called a cyclic factor or versor. In particular the number $i = \sqrt{-1}$ will effect a rotation through 90° when used as a multiplier and is known as a quadrantal versor. The series of powers i, $i^2 = -1$, $i^3 = -i$, $i^4 = 1$ give rotations through 90°, 180°, 270°, 360°. This fact is often given as the reason for laying off pure imaginary numbers bi along an axis at right angles to the axis of reals.

As a particular product, the nth power of a complex number is

$$\alpha^n = (a + ib)^n = [r(\cos\phi + i\sin\phi)]^n = r^n(\cos n\phi + i\sin n\phi); \quad (15)$$

and

$$(\cos\phi + i\sin\phi)^n = \cos n\phi + i\sin n\phi, \quad (15')$$

which is a special case, is known as *De Moivre's Theorem* and is of use in evaluating the functions of $n\phi$; for the binomial theorem may be applied and the real and imaginary parts of the expansion may be equated to $\cos n\phi$ and $\sin n\phi$. Hence

$$\cos n\phi = \cos^n\phi - \frac{n(n-1)}{2!}\cos^{n-2}\phi\sin^2\phi + \frac{n(n-1)(n-2)(n-3)}{4!}\cos^{n-4}\phi\sin^4\phi - \cdots \quad (16)$$

$$\sin n\phi = n\cos^{n-1}\phi\sin\phi - \frac{n(n-1)(n-2)}{3!}\cos^{n-3}\phi\sin^3\phi + \cdots.$$

As the nth root $\sqrt[n]{\alpha}$ of α must be a number which when raised to the nth power gives α, the nth root may be written as

$$\sqrt[n]{\alpha} = \sqrt[n]{r}(\cos\phi/n + i\sin\phi/n). \quad (17)$$

The angle ϕ, however, may have any of the set of values

$$\phi, \quad \phi + 2\pi, \quad \phi + 4\pi, \quad \cdots, \quad \phi + 2(n-1)\pi,$$

and the nth parts of these give the n different angles

$$\frac{\phi}{n}, \quad \frac{\phi}{n}+\frac{2\pi}{n}, \quad \frac{\phi}{n}+\frac{4\pi}{n}, \quad \cdots, \quad \frac{\phi}{n}+\frac{2(n-1)\pi}{n}. \tag{18}$$

Hence there may be found just n different nth roots of any given complex number (including, of course, the reals).

The *roots of unity* deserve mention. The equation $x^n = 1$ has in the real domain one or two roots according as n is odd or even. But if 1 be regarded as a complex number of which the pure imaginary part is zero, it may be represented by a point at a unit distance from the origin upon the axis of reals; the magnitude of 1 is 1 and the angle of 1 is $0, 2\pi, \cdots, 2(n-1)\pi$. The nth roots of 1 will therefore have the magnitude 1 and one of the angles $0, 2\pi/n, \cdots, 2(n-1)\pi/n$. The n nth roots are therefore

$$1, \quad \alpha = \cos\frac{2\pi}{n}+i\sin\frac{2\pi}{n}, \quad \alpha^2 = \cos\frac{4\pi}{n}+i\sin\frac{4\pi}{n}, \quad \cdots,$$
$$\alpha^{n-1} = \cos\frac{2(n-1)\pi}{n}+i\sin\frac{2(n-1)\pi}{n},$$

and may be evaluated with a table of natural functions. Now $x^n - 1 = 0$ is factorable as $(x-1)(x^{n-1}+x^{n-2}+\cdots+x+1)=0$, and it therefore follows that the nth roots other than 1 must all satisfy the equation formed by setting the second factor equal to 0. As α in particular satisfies this equation and the other roots are $\alpha^2, \cdots, \alpha^{n-1}$, it follows that the sum of the n nth roots of unity is zero.

EXERCISES

1. Prove the distributive law of multiplication for complex numbers.

2. By definition the pair of imaginaries $a+bi$ and $a-bi$ are called *conjugate imaginaries*. Prove that (α) the sum and the product of two conjugate imaginaries are real; and conversely (β) if the sum and the product of two imaginaries are both real, the imaginaries are conjugate.

3. Show that if $P(x, y)$ is a symmetric polynomial in x and y with real coefficients so that $P(x, y) = P(y, x)$, then if conjugate imaginaries be substituted for x and y, the value of the polynomial will be real.

4. Show that if $a+bi$ is a root of an algebraic equation $P(x)=0$ with real coefficients, then $a-bi$ is also a root of the equation.

5. Carry out the indicated operations algebraically and make a graphical representation for every number concerned and for the answer:

(α) $(1+i)^3$, (β) $(1+\sqrt{3}\,i)(1-i)$, (γ) $(3+\sqrt{-2})(4+\sqrt{-5})$,

(δ) $\dfrac{1+i}{1-i}$, (ϵ) $\dfrac{1+i\sqrt{3}}{1-i\sqrt{3}}$, ($\zeta$) $\dfrac{5}{\sqrt{2}-i\sqrt{3}}$,

(η) $\dfrac{(1-i)^2}{(1+i)^3}$, (θ) $\dfrac{1}{(1+i)^2}+\dfrac{1}{(1-i)^2}$, ($\iota$) $\left(\dfrac{-1+\sqrt{-3}}{2}\right)^3$.

6. Plot and find the modulus and angle in the following cases:

(α) -2, (β) $-2\sqrt{-1}$, (γ) $3+4i$, (δ) $\frac{1}{2}-\frac{1}{2}\sqrt{-3}$.

7. Show that *the modulus of a quotient of two numbers is the quotient of the moduli and that the angle is the angle of the numerator less that of the denominator.*

8. Carry out the indicated operations trigonometrically and plot:

(α) The examples of Ex. 5, (β) $\sqrt{1+i}\sqrt{1-i}$, (γ) $\sqrt{-2+2\sqrt{3}i}$,

(δ) $(\sqrt{1+i}+\sqrt{1-i})^2$, ($\epsilon$) $\sqrt{\sqrt{2}+\sqrt{-2}}$, ($\zeta$) $\sqrt[3]{2+2\sqrt{3}i}$,

(η) $\sqrt[4]{16(\cos 200^\circ + i \sin 200^\circ)}$, ($\theta$) $\sqrt[5]{-1}$, (ι) $\sqrt[6]{8i}$.

9. Find the equations of analytic geometry which represent the transformation equivalent to multiplication by $\alpha = -1+\sqrt{-3}$.

10. Show that $|z-\alpha| = r$, where z is a variable and α a fixed complex number, is the equation of the circle $(x-a)^2+(y-b)^2 = r^2$.

11. Find $\cos 5x$ and $\cos 8x$ in terms of $\cos x$, and $\sin 6x$ and $\sin 7x$ in terms of $\sin x$.

12. Obtain to four decimal places the five roots $\sqrt[5]{1}$.

13. If $z = x+iy$ and $z' = x'+iy'$, show that $z' = (\cos\phi - i\sin\phi)z - \alpha$ is the formula for shifting the axes through the vector distance $\alpha = a+ib$ to the new origin (a, b) and turning them through the angle ϕ. Deduce the ordinary equations of transformation.

14. Show that $|z-\alpha| = k|z-\beta|$, where k is real, is the equation of a circle; specify the position of the circle carefully. Use the theorem: The locus of points whose distances to two fixed points are in a constant ratio is a circle the diameter of which is divided internally and externally in the same ratio by the fixed points.

15. The transformation $z' = \dfrac{az+b}{cz+d}$, where a, b, c, d are complex and $ad-bc \neq 0$, is called the *general linear transformation* of z into z'. Show that

$$|z'-\alpha'| = k|z'-\beta'| \quad \text{becomes} \quad |z-\alpha| = k\left|\frac{c\alpha+d}{c\beta+d}\right|\cdot|z-\beta|.$$

Hence infer that the transformation carries circles into circles, and points which divide a diameter internally and externally in the same ratio into points which divide some diameter of the new circle similarly, but generally with a different ratio.

73. Functions of a complex variable. Let $z = x+iy$ be a complex variable representable geometrically as a variable point in the xy-plane, which may be called the *complex plane*. As z determines the two real numbers x and y, any function $F(x, y)$ which is the sum of two single valued real functions in the form

$$F(x, y) = X(x, y) + iY(x, y) = R(\cos\Phi + i\sin\Phi) \tag{19}$$

will be completely determined in value if z is given. Such a function is called a *complex function* (and not a function of the complex variable, for reasons that will appear later). The magnitude and angle of the function are determined by

$$R = \sqrt{X^2+Y^2}, \qquad \cos\Phi = \frac{X}{R}, \ \sin\Phi = \frac{Y}{R}. \tag{20}$$

The function F is continuous by definition when and only when both X and Y are continuous functions of (x, y); R is then continuous in (x, y) and F can vanish only when $R = 0$; the angle Φ regarded as a function of (x, y) is also continuous and determinate (except for the additive $2n\pi$) unless $R = 0$, in which case X and Y also vanish and the expression for Φ involves an indeterminate form in two variables and is generally neither determinate nor continuous (§ 44).

If the derivative of F with respect to z were sought for the value $z = a + ib$, the procedure would be entirely analogous to that in the case of a real function of a real variable. The increment $\Delta z = \Delta x + i\Delta y$ would be assumed for z and ΔF would be computed and the quotient $\Delta F/\Delta z$ would be formed. Thus by the Theorem of the Mean (§ 46),

$$\frac{\Delta F}{\Delta z} = \frac{\Delta X + i\Delta Y}{\Delta x + i\Delta y} = \frac{(X'_x + iY'_x)\Delta x + (X'_y + iY'_y)\Delta y}{\Delta x + i\Delta y} + \zeta, \qquad (21)$$

where the derivatives are formed for (a, b) and where ζ is an infinitesimal complex number. When Δz approaches 0, both Δx and Δy must approach 0 without any implied relation between them. In general the limit of $\Delta F/\Delta z$ is a double limit (§ 44) and may therefore depend on the way in which Δx and Δy approach their limit 0.

Now if first $\Delta y \doteq 0$ and then subsequently $\Delta x \doteq 0$, the value of the limit of $\Delta F/\Delta z$ is $X'_x + iY'_x$ taken at the point (a, b); whereas if first $\Delta x \doteq 0$ and then $\Delta y \doteq 0$, the value is $-iX'_y + Y'_y$. Hence if the limit of $\Delta F/\Delta z$ is to be independent of the way in which Δz approaches 0, it is surely necessary that

$$\frac{\partial X}{\partial x} + i\frac{\partial Y}{\partial x} = -i\frac{\partial X}{\partial y} + \frac{\partial Y}{\partial y},$$

or
$$\frac{\partial X}{\partial x} = \frac{\partial Y}{\partial y} \quad \text{and} \quad \frac{\partial X}{\partial y} = -\frac{\partial Y}{\partial x}. \qquad (22)$$

And conversely if these relations are satisfied, then

$$\frac{\Delta F}{\Delta z} = \left(\frac{\partial X}{\partial x} + i\frac{\partial Y}{\partial x}\right) + \zeta = \left(\frac{\partial Y}{\partial y} - i\frac{\partial X}{\partial y}\right) + \zeta;$$

and the limit is $X'_x + iY'_x = Y'_y - iX'_y$ taken at the point (a, b), and is independent of the way in which Δz approaches zero. The desirability of having at least the ordinary functions differentiable suggests the definition: *A complex function $F(x, y) = X(x, y) + iY(x, y)$ is considered as a function of the complex variable $z = x + iy$ when and only when X and Y are in general differentiable and satisfy the relations (22).* In this case *the derivative is*

$$F'(z) = \frac{dF}{dz} = \frac{\partial X}{\partial x} + i\frac{\partial Y}{\partial x} = \frac{\partial Y}{\partial y} - i\frac{\partial X}{\partial y}. \tag{23}$$

These conditions may also be expressed in polar coördinates (Ex. 2).

A few words about the function $\Phi(x, y)$. This is a multiple valued function of the variables (x, y), and the difference between two neighboring branches is the constant 2π. The application of the discussion of § 45 to this case shows at once that, in any simply connected region of the complex plane which contains no point (a, b) such that $R(a, b) = 0$, the different branches of $\Phi(x, y)$ may be entirely separated so that the value of Φ must return to its initial value when any closed curve is described by the point (x, y). If, however, the region is multiply connected or contains points for which $R = 0$ (which makes the region multiply connected because these points must be cut out), it may happen that there will be circuits for which Φ, although changing continuously, will not return to its initial value. Indeed if it can be shown that Φ does not return to its initial value when changing continuously as (x, y) describes the boundary of a region simply connected except for the excised points, it may be inferred that there must be points in the region for which $R = 0$.

An application of these results may be made to give a very simple demonstration of *the fundamental theorem of algebra that every equation of the nth degree has at least one root.* Consider the function

$$F(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n = X(x, y) + iY(x, y),$$

where X and Y are found by writing z as $x + iy$ and expanding and rearranging. The functions X and Y will be polynomials in (x, y) and will therefore be everywhere finite and continuous in (x, y). Consider the angle Φ of F. Then

$$\Phi = \text{ang. of } F = \text{ang. of } z^n\left(1 + \frac{a_1}{z} + \cdots + \frac{a_{n-1}}{z^{n-1}} + \frac{a_n}{z^n}\right) = \text{ang. of } z^n + \text{ang. of } (1 + \cdots).$$

Next draw about the origin a circle of radius r so large that

$$\left|\frac{a_1}{z}\right| + \cdots + \left|\frac{a_{n-1}}{z^{n-1}}\right| + \left|\frac{a_n}{z^n}\right| = \frac{|a_1|}{r} + \cdots + \frac{|a_{n-1}|}{r^{n-1}} + \frac{|a_n|}{r^n} < \epsilon.$$

Then for all points z upon the circumference the angle of F is

$$\Phi = \text{ang. of } F = n\,(\text{ang. of } z) + \text{ang. of } (1 + \eta), \qquad |\eta| < \epsilon.$$

Now let the point (x, y) describe the circumference. The angle of z will change by 2π for the complete circuit. Hence Φ must change by $2n\pi$ and does not return to its initial value. Hence there is within the circle at least one point (a, b) for which $R(a, b) = 0$ and consequently for which $X(a, b) = 0$ and $Y(a, b) = 0$ and $F(a, b) = 0$. Thus if $\alpha = a + ib$, then $F(\alpha) = 0$ and the equation $F(z) = 0$ is seen to have at least the one root α. It follows that $z - \alpha$ is a factor of $F(z)$; and hence by induction it may be seen that $F(z) = 0$ has just n roots.

74. The discussion of the algebra of complex numbers showed how the sum, difference, product, quotient, real powers, and real roots of such numbers could be found, and hence made it possible to compute the value of any given algebraic expression or function of z for a given value of z. It remains to show that any algebraic expression in z is

really a function of z in the sense that it has a derivative with respect to z, and to find the derivative. Now the differentiation of an algebraic function of the variable x was made to depend upon the formulas of differentiation, (6) and (7) of § 2. A glance at the methods of derivation of these formulas shows that they were proved by ordinary algebraic manipulations such as have been seen to be equally possible with imaginaries as with reals. It therefore may be concluded that *an algebraic expression in z has a derivative with respect to z and that derivative may be found just as if z were a real variable.*

The case of the elementary functions e^z, $\log z$, $\sin z$, $\cos z$, $\cdots$ other than algebraic is different; for these functions have not been defined for complex variables. Now in seeking to define these functions when z is complex, an effort should be made to define in such a way that: 1° when z is real, the new and the old definitions become identical; and 2° the rules of operation with the function shall be as nearly as possible the same for the complex domain as for the real. Thus it would be desirable that $De^z = e^z$ and $e^{z+w} = e^z e^w$, when z and w are complex. With these ideas in mind one may proceed to define the elementary functions for complex arguments. Let

$$e^z = R(x, y)\left[\cos\Phi(x, y) + i\sin\Phi(x, y)\right]. \tag{24}$$

The derivative of this function is, by the first rule of (23),

$$\begin{aligned} De^z &= \frac{\partial}{\partial x}(R\cos\Phi) + i\frac{\partial}{\partial x}(R\sin\Phi) \\ &= (R'_x\cos\Phi - R\sin\Phi\cdot\Phi'_x) + i(R'_x\sin\Phi + R\cos\Phi\cdot\Phi'_x), \end{aligned}$$

and if this is to be identical with e^z above, the equations

$$\begin{matrix} R'_x\cos\Phi - R\Phi'_x\sin\Phi = R\cos\Phi \\ R'_x\sin\Phi + R\Phi'_x\cos\Phi = R\sin\Phi \end{matrix} \quad \text{or} \quad \begin{matrix} R'_x = R \\ \Phi'_x = 0 \end{matrix}$$

must hold, where the second pair is obtained by solving the first. If the second form of the derivative in (23) had been used, the results would have been $R'_y = 0$, $\Phi'_y = 1$. It therefore appears that if the derivative of e^z, however computed, is to be e^z, then

$$R'_x = R, \quad R'_y = 0, \quad \Phi'_x = 0, \quad \Phi'_y = 1$$

are four conditions imposed upon R and Φ. These conditions will be satisfied if $R = e^x$ and $\Phi = y$.* Hence define

$$e^z = e^{x+iy} = e^x(\cos y + i\sin y). \tag{25}$$

* The use of the more general solutions $R = Ce^x$, $\Phi = y + C'$ would lead to expressions which would not reduce to e^x when $y = 0$ and $z = x$ or would not satisfy $e^{z+w} = e^z e^w$.

With this definition De^z is surely e^z, and it is readily shown that the exponential law $e^{z+w} = e^z e^w$ holds.

For the special values $\frac{1}{2}\pi i$, πi, $2\pi i$ of z the value of e^z is

$$e^{\frac{1}{2}\pi i} = i, \quad e^{\pi i} = -1, \quad e^{2\pi i} = 1.$$

Hence it appears that if $2n\pi i$ be added to z, e^z is unchanged;

$$e^{z+2n\pi i} = e^z, \quad \text{period} \quad 2\pi i. \tag{26}$$

Thus *in the complex domain e^z has the period $2\pi i$*, just as $\cos x$ and $\sin x$ have the real period 2π. This relation is inherent; for

$$e^{yi} = \cos y + i \sin y, \quad e^{-yi} = \cos y - i \sin y,$$

and
$$\cos y = \frac{e^{yi} + e^{-yi}}{2}, \quad \sin y = \frac{e^{yi} - e^{-yi}}{2i}. \tag{27}$$

The trigonometric functions of a real variable y may be expressed in terms of the exponentials of yi and $-yi$. As the exponential has been defined for all complex values of z, it is natural to use (27) to define the trigonometric functions for complex values as

$$\cos z = \frac{e^{zi} + e^{-zi}}{2}, \quad \sin z = \frac{e^{zi} - e^{-zi}}{2i}. \tag{27'}$$

With these definitions the ordinary formulas for $\cos(z + w)$, $D \sin z$, $\cdots$ may be obtained and be seen to hold for complex arguments, just as the corresponding formulas were derived for the hyperbolic functions (§ 5).

As in the case of reals, the logarithm $\log z$ will be defined for complex numbers as the inverse of the exponential. Thus

$$\text{if} \quad e^z = w, \quad \text{then} \quad \log w = z + 2n\pi i, \tag{28}$$

where the periodicity of the function e^z shows that *the logarithm is not uniquely determined but admits the addition of $2n\pi i$ to any one of its values*, just as $\tan^{-1} x$ admits the addition of $n\pi$. If w is written as a complex number $u + iv$ with modulus $r = \sqrt{u^2 + v^2}$ and with the angle ϕ, it follows that

$$w = u + iv = r(\cos\phi + i \sin\phi) = re^{\phi i} = e^{\log r + \phi i}; \tag{29}$$

and
$$\log w = \log r + \phi i = \log\sqrt{u^2 + v^2} + i \tan^{-1}(v/u)$$

is the expression for the logarithm of w in terms of the modulus and angle of w; the $2n\pi i$ may be added if desired.

To this point the expression of a power a^b, where the exponent b is imaginary, has had no definition. The definition may now be given in terms of exponentials and logarithms. Let

$$a^b = e^{b \log a} \quad \text{or} \quad \log a^b = b \log a.$$

In this way the problem of computing a^b is reduced to one already solved. From the very definition it is seen that the logarithm of a power is the product of the exponent by the logarithm of the base, as in the case of reals. To indicate the path that has been followed in defining functions, a sort of family tree may be made.

real numbers, x — real powers and roots of reals, x^n

real angles, x — real trigonometric functions, $\cos x$, $\sin x$, $\tan^{-1}x$, $\cdots$

exponentials, logarithms of reals, e^x, $\log x$

real powers and roots of imaginaries, z^n

exponentials of imaginaries, e^z

logarithms of imaginaries, $\log z$

imaginary powers, z^α

trigonometric functions of imaginaries

EXERCISES

1. Show that the following complex functions satisfy the conditions (22) and are therefore functions of the complex variable z. Find $F'(z)$:

(α) $x^2 - y^2 + 2\,ixy$. $\qquad (\beta)$ $x^3 - 3(xy^2 + x^2 - y^2) + i(3x^2y - y^3 - 6xy)$.

(γ) $\dfrac{x}{x^2+y^2} - i\dfrac{y}{x^2+y^2}$, $\qquad (\delta)$ $\log\sqrt{x^2+y^2} + i\tan^{-1}\dfrac{y}{x}$.

(ϵ) $e^x\cos y + ie^x\sin y$, $\qquad (\zeta)$ $\sin x\sinh y + i\cos x\cosh y$.

2. Show that in polar coördinates the conditions for the existence of $F'(z)$ are

$$\frac{\partial X}{\partial r} = \frac{1}{r}\frac{\partial Y}{\partial \phi},\quad \frac{\partial Y}{\partial r} = -\frac{1}{r}\frac{\partial X}{\partial \phi}\quad \text{with}\quad F'(z) = \left(\frac{\partial X}{\partial r} + i\frac{\partial Y}{\partial r}\right)(\cos\phi - i\sin\phi).$$

3. Use the conditions of Ex. 2 to show from $D\log z = z^{-1}$ that $\log z = \log r + \phi i$.

4. From the definitions given above prove the formulas

$$(\alpha)\quad \sin(x+iy) = \sin x\cosh y + i\cos x\sinh y.$$

$$(\beta)\quad \cos(x+iy) = \cos x\cosh y - i\sin x\sinh y.$$

$$(\gamma)\quad \tan(x+iy) = \frac{\sin 2x + i\sinh 2y}{\cos 2x + \cosh 2y}.$$

5. Find to three decimals the complex numbers which express the values of:

(α) $e^{\frac{1}{4}\pi i}$. $\quad(\beta)$ e^i. $\quad(\gamma)$ $e^{\frac{1}{2}+\frac{1}{2}\sqrt{-3}}$. $\quad(\delta)$ e^{-1-i}.

(ϵ) $\sin\frac{1}{4}\pi i$. $\quad(\zeta)$ $\cos i$. $\quad(\eta)$ $\sin(\frac{1}{2}+\frac{1}{2}\sqrt{-3})$. $\quad(\theta)$ $\tan(-1-i)$.

(ι) $\log(-1)$. $\quad(\kappa)$ $\log i$. $\quad(\lambda)$ $\log(\frac{1}{2}+\frac{1}{2}\sqrt{-3})$. $\quad(\mu)$ $\log(-1-i)$.

6. Owing to the fact that $\log a$ is multiple valued, a^b is multiple valued in such a manner that any one value may be multiplied by $e^{2n\pi bi}$. Find one value of each of the following and several values of one of them:

(α) 2^i. $\quad(\beta)$ i^i. $\quad(\gamma)$ $\sqrt[i]{i}$. $\quad(\delta)$ $\sqrt[i]{2}$. $\quad(\epsilon)$ $(\frac{1}{2}+\frac{1}{2}\sqrt{-3})^{\frac{3}{\pi}i+1}$.

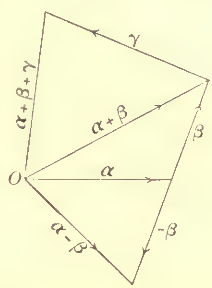
7. Show that $Da^z = a^z \log a$ when a and z are complex.

8. Show that $(a^b)^c = a^{bc}$; and fill in such other steps as may be suggested by the work in the text, which for the most part has merely been sketched in a broad way.

9. Show that if $f(z)$ and $g(z)$ are two functions of a complex variable, then $f(z) \pm g(z)$, $\alpha f(z)$ with α a complex constant, $f(z)g(z)$, $f(z)/g(z)$ are also functions of z.

10. Obtain logarithmic expressions for the inverse trigonometric functions. Find $\sin^{-1} i$.

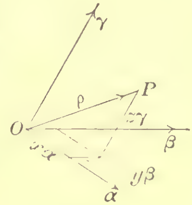
75. Vector sums and products. As stated in § 71, a vector is a quantity which has magnitude and direction. If the magnitudes of two vectors are equal and the directions of the two vectors are the same, the vectors are said to be equal irrespective of the position which they occupy in space. The vector $-\alpha$ is by definition a vector which has the same magnitude as α but the opposite direction. The vector $m\alpha$ is a vector which has the same direction as α (or the opposite) and is m (or $-m$) times as long. The law of vector or geometric addition is the parallelogram or triangle law (§ 71) and is still applicable when the vectors do not lie in a plane but have any directions in space; for any two vectors brought end to end determine a plane in which the construction may be carried out. Vectors will be designated by Greek small letters or by letters in heavy type. The relations of equality or similarity between triangles establish the rules



$$\alpha + \beta = \beta + \alpha, \quad \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma, \quad m(\alpha + \beta) = m\alpha + m\beta \quad (30)$$

as true for vectors as well as for numbers whether real or complex. A vector is said to be zero when its magnitude is zero, and it is written 0. From the definition of addition it follows that $\alpha + 0 = \alpha$. In fact *as far as addition, subtraction, and multiplication by numbers are concerned, vectors obey the same formal laws as numbers.*

A vector ρ may be resolved into components parallel to any three given vectors α, β, γ which are not parallel to any one plane. For let a parallelepiped be constructed with its edges parallel to the three given vectors and with its diagonal equal to the vector whose components are desired. The edges of the parallelepiped are then certain



multiples $x\alpha$, $y\beta$, $z\gamma$ of α, β, γ; and these are the desired components of ρ. The vector ρ may be written as

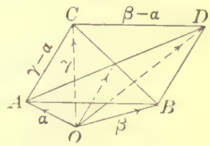
$$\rho = x\alpha + y\beta + z\gamma.^* \tag{31}$$

It is clear that two equal vectors would necessarily have the same components along three given directions and that the components of a zero vector would all be zero. Just as the equality of two complex numbers involved the two equalities of the respective real and imaginary parts, so the equality of two vectors as

$$\rho = x\alpha + y\beta + z\gamma = x'\alpha + y'\beta + z'\gamma = \rho' \tag{31'}$$

involves the three equations $x = x'$, $y = y'$, $z = z'$.

As a problem in the use of vectors let there be given the three vectors α, β, γ from an assumed origin O to three vertices of a parallelogram; required the vector to the other vertex, the vector expressions for the sides and diagonals of the parallelogram, and the proof of the fact that the diagonals bisect each other. Consider the figure. The side AB is, by the triangle law, that vector which when added to $OA = \alpha$ gives $OB = \beta$, and hence it must be that $AB = \beta - \alpha$. In like manner $AC = \gamma - \alpha$. Now OD is the sum of OC and CD, and $CD = AB$; hence $OD = \gamma + \beta - \alpha$. The diagonal AD is the difference of the vectors OD and OA, and is therefore $\gamma + \beta - 2\alpha$. The diagonal BC is $\gamma - \beta$. Now the vector from O to the middle point of BC may be found by adding to OB one half of BC. Hence this vector is $\beta + \frac{1}{2}(\gamma - \beta)$ or $\frac{1}{2}(\beta + \gamma)$. In like manner the vector to the middle point of AD is seen to be $\alpha + \frac{1}{2}(\gamma + \beta - 2\alpha)$ or $\frac{1}{2}(\gamma + \beta)$, which is identical with the former. The two middle points therefore coincide and the diagonals bisect each other.



Let α and β be any two vectors, $|\alpha|$ and $|\beta|$ their respective lengths, and $\angle(\alpha, \beta)$ the angle between them. For convenience the vectors may be considered to be laid off from the same origin. The product of the lengths of the vectors by the cosine of the angle between the vectors is called the *scalar product*,

$$\text{scalar product} = \alpha \cdot \beta = |\alpha||\beta| \cos \angle(\alpha, \beta), \tag{32}$$

of the two vectors and is denoted by placing a dot between the letters. This combination, called the scalar product, is a number, not a vector. As $|\beta| \cos \angle(\alpha, \beta)$ is the projection of β upon the direction of α, the scalar product may be stated to be equal to the product of the length of either vector by the length of the projection of the other upon it. In particular if either vector were of unit length, the scalar product would be the projection of the other upon it, with proper regard for

* The numbers x, y, z are the oblique coördinates of the terminal end of ρ (if the initial end be at the origin) referred to a set of axes which are parallel to α, β, γ and upon which the unit lengths are taken as the lengths of α, β, γ respectively.

the sign; and if both vectors are unit vectors, the product is the cosine of the angle between them.

The scalar product, from its definition, is *commutative* so that $\alpha\cdot\beta=\beta\cdot\alpha$. Moreover $(m\alpha)\cdot\beta = \alpha\cdot(m\beta) = m\,(\alpha\cdot\beta)$, thus allowing a numerical factor m to be combined with either factor of the product. Furthermore the *distributive law*

$$\alpha\cdot(\beta+\gamma) = \alpha\cdot\beta + \alpha\cdot\gamma \quad \text{or} \quad (\alpha+\beta)\cdot\gamma = \alpha\cdot\gamma + \beta\cdot\gamma \tag{33}$$

is satisfied as in the case of numbers. For if α be written as the product $a\alpha_1$ of its length a by a vector α_1 of unit length in the direction of α, the first equation becomes

$$a\alpha_1\cdot(\beta+\gamma) = a\alpha_1\cdot\beta + a\alpha_1\cdot\gamma \quad \text{or} \quad \alpha_1\cdot(\beta+\gamma) = \alpha_1\cdot\beta + \alpha_1\cdot\gamma.$$

And now $\alpha_1\cdot(\beta+\gamma)$ is the projection of the sum $\beta+\gamma$ upon the direction of α, and $\alpha_1\cdot\beta + \alpha_1\cdot\gamma$ is the sum of the projections of β and γ upon this direction; by the law of projections these are equal and hence the distributive law is proved.

The associative law does not hold for scalar products; for $(\alpha\cdot\beta)\gamma$ means that the vector γ is multiplied by the number $\alpha\cdot\beta$, whereas $\alpha(\beta\cdot\gamma)$ means that α is multiplied by $(\beta\cdot\gamma)$, a very different matter. The laws of cancellation cannot hold: for if

$$\alpha\cdot\beta = 0, \quad \text{then} \quad |\alpha|\,|\beta|\cos\angle(\alpha,\beta) = 0, \tag{34}$$

and the vanishing of the scalar product $\alpha\cdot\beta$ implies either that one of the factors is 0 or that the two vectors are perpendicular. In fact $\alpha\cdot\beta=0$ is called the *condition of perpendicularity*. It should be noted, however, that if a vector ρ satisfies

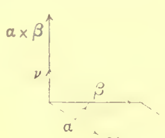
$$\rho\cdot\alpha = 0, \qquad \rho\cdot\beta = 0, \qquad \rho\cdot\gamma = 0, \tag{35}$$

three conditions of perpendicularity with three vectors α, β, γ not parallel to the same plane, the inference is that $\rho = 0$.

76. Another product of two vectors is the *vector product*,

$$\text{vector product} = \alpha\times\beta = \nu\,|\alpha|\,|\beta|\sin\angle(\alpha,\beta), \tag{36}$$

where ν represents a vector of unit length normal to the plane of α and β upon that side on which rotation from α to β through an angle of less than 180° appears positive or counterclockwise. Thus the vector product is itself a vector of which the direction is perpendicular to each factor, and of which the magnitude is the product of the magnitudes into the sine of the included angle. The magnitude is therefore equal to the area of the parallelogram of which the vectors α and β are the sides.



The vector product will be represented by a cross inserted between the letters.

As rotation from β to α is the opposite of that from α to β, it follows from the definition of the vector product that

$$\beta\times\alpha = -\alpha\times\beta, \quad \text{not} \quad \alpha\times\beta = \beta\times\alpha, \tag{37}$$

and the product is *not commutative*, the order of the factors must be carefully observed. Furthermore the equation

$$\alpha\times\beta = \nu\, \alpha\, \beta \sin\angle(\alpha, \beta) = 0 \tag{38}$$

implies either that one of the factors vanishes or that the vectors α and β are parallel. Indeed the condition $\alpha\times\beta = 0$ is called the *condition of parallelism*. The laws of cancellation do not hold. The associative law also does not hold: for $(\alpha\times\beta)\times\gamma$ is a vector perpendicular to $\alpha\times\beta$ and γ, and since $\alpha\times\beta$ is perpendicular to the plane of α and β, the vector $(\alpha\times\beta)\times\gamma$ perpendicular to it must lie in the plane of α and β; whereas the vector $\alpha\times(\beta\times\gamma)$, by similar reasoning, must lie in the plane of β and γ; and hence the two vectors cannot be equal except in the very special case where each was parallel to β which is common to the two planes.

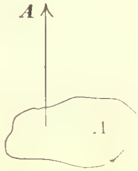
But the operation $(m\alpha)\times\beta = \alpha\times(m\beta) = m(\alpha\times\beta)$, which consists in allowing the transference of a numerical factor to any position in the product, does hold; and so does the *distributive law*

$$\alpha\times(\beta+\gamma) = \alpha\times\beta + \alpha\times\gamma \quad \text{and} \quad (\alpha+\beta)\times\gamma = \alpha\times\gamma + \beta\times\gamma, \tag{39}$$

the proof of which will be given below. In expanding according to the distributive law care must be exercised to keep the order of the factors in each vector product the same on both sides of the equation, owing to the failure of the commutative law; an interchange of the order of the factors changes the sign. It might seem as if any algebraic operations where so many of the laws of elementary algebra fail as in the case of vector products would be too restricted to be very useful; that this is not so is due to the astonishingly great number of problems in which the analysis can be carried on with only the laws of addition and the distributive law of multiplication combined with the possibility of transferring a numerical factor from one position to another in a product; in addition to these laws, the scalar product $\alpha\cdot\beta$ is commutative and the vector product $\alpha\times\beta$ is commutative except for change of sign.

In addition to segments of lines, *plane areas may be regarded as vector quantities*: for a plane area has magnitude (the amount of the area) and direction (the direction of the normal to its plane). To specify on which side of the plane the normal lies, some convention must be made. If the area is part of a surface inclosing a portion of space, the

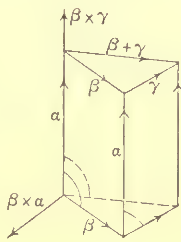
normal is taken as the exterior normal. If the area lies in an isolated plane, its positive side is determined only in connection with some assigned direction of description of its bounding curve; the rule is: If a person is assumed to walk along the boundary of an area in an assigned direction and upon that side of the plane which causes the inclosed area to lie upon his left, he is said to be upon the positive side (for the assigned direction of description of the boundary), and the vector which represents the area is the normal to that side. It has been mentioned that the vector product represented an area.



That the projection of a plane area upon a given plane gives an area which is the original area multiplied by the cosine of the angle between the two planes is a fundamental fact of projection, following from the simple fact that lines parallel to the intersection of the two planes are unchanged in length whereas lines perpendicular to the intersection are multiplied by the cosine of the angle between the planes. As the angle between the normals is the same as that between the planes, *the projection of an area upon a plane and the projection of the vector representing the area upon the normal to the plane are equivalent.* The projection of a *closed* area upon a plane is zero; for the area in the projection is covered twice (or an even number of times) with opposite signs and the total algebraic sum is therefore 0.

To prove the law $\alpha\times(\beta+\gamma)=\alpha\times\beta+\alpha\times\gamma$ and illustrate the use of the vector interpretation of areas, construct a triangular prism with the triangle on β, γ, and $\beta+\gamma$ as base and α as lateral edge. The total vector expression for the surface of this prism is

$$\beta\times\alpha+\gamma\times\alpha+\alpha\times(\beta+\gamma)+\tfrac{1}{2}(\beta\times\gamma)-\tfrac{1}{2}\beta\times\gamma=0,$$



and vanishes because the surface is closed. A cancellation of the equal and opposite terms (the two bases) and a simple transposition combined with the rule $\beta\times\alpha=-\alpha\times\beta$ gives the result

$$\alpha\times(\beta+\gamma)=-\beta\times\alpha-\gamma\times\alpha=\alpha\times\beta+\alpha\times\gamma.$$

A system of *vectors of reference* which is particularly useful consists of three vectors **i**, **j**, **k** of unit length directed along the axes X, Y, Z drawn so that rotation from X to Y appears positive from the side of the xy-plane upon which Z lies. The components of any vector **r** drawn from the origin to the point (x, y, z) are

$$x\mathbf{i},\quad y\mathbf{j},\quad z\mathbf{k},\quad \text{and}\quad \mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}.$$

The products of **i**, **j**, **k** into each other are, from the definitions,

$$\begin{gathered}\mathbf{i}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1,\\ \mathbf{i}\cdot\mathbf{j} = \mathbf{j}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{k} = \mathbf{k}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{i} = \mathbf{i}\cdot\mathbf{k} = 0,\\ \mathbf{i}\times\mathbf{i} = \mathbf{j}\times\mathbf{j} = \mathbf{k}\times\mathbf{k} = 0,\\ \mathbf{i}\times\mathbf{j} = -\mathbf{j}\times\mathbf{i} = \mathbf{k}, \quad \mathbf{j}\times\mathbf{k} = -\mathbf{k}\times\mathbf{j} = \mathbf{i}, \quad \mathbf{k}\times\mathbf{i} = -\mathbf{i}\times\mathbf{k} = \mathbf{j}.\end{gathered} \tag{40}$$

By means of these products and the distributive laws for scalar and vector products, any given products may be expanded. Thus if

$$\alpha = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \quad \text{and} \quad \beta = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k},$$

then

$$\begin{gathered}\alpha\cdot\beta = a_1b_1 + a_2b_2 + a_3b_3,\\ \alpha\times\beta = (a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k},\end{gathered} \tag{41}$$

by direct multiplication. In this way a passage may be made from vector formulas to Cartesian formulas whenever desired.

EXERCISES

1. Prove geometrically that $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ and $m(\alpha + \beta) = m\alpha + m\beta$.

2. If α and β are the vectors from an assumed origin to A and B and if C divides AB in the ratio $m:n$, show that the vector to C is $\gamma = (n\alpha + m\beta)/(m+n)$.

3. In the parallelogram $ABCD$ show that the line BE connecting the vertex to the middle point of the opposite side CD is trisected by the diagonal AD and trisects it.

4. Show that the medians of a triangle meet in a point and are trisected.

5. If m_1 and m_2 are two masses situated at P_1 and P_2, the *center of gravity* or *center of mass* of m_1 and m_2 is defined as that point G on the line P_1P_2 which divides P_1P_2 inversely as the masses. Moreover if G_1 is the center of mass of a number of masses of which the total mass is M_1 and if G_2 is the center of mass of a number of other masses whose total mass is M_2, the same rule applied to M_1 and M_2 and G_1 and G_2 gives the center of gravity G of the total number of masses. Show that

$$\mathbf{r} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} \quad \text{and} \quad \mathbf{r} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_n\mathbf{r}_n}{m_1 + m_2 + \cdots + m_n} = \frac{\Sigma m\mathbf{r}}{\Sigma m},$$

where **r** denotes the vector to the center of gravity. Resolve into components to show

$$\bar{x} = \frac{\Sigma mx}{\Sigma m}, \quad \bar{y} = \frac{\Sigma my}{\Sigma m}, \quad \bar{z} = \frac{\Sigma mz}{\Sigma m}.$$

6. If α and β are two fixed vectors and ρ a variable vector, all being laid off from the same origin, show that $(\rho - \beta)\cdot\alpha = 0$ is the equation of a plane through the end of β perpendicular to α.

7. Let α, β, γ be the vectors to the vertices A, B, C of a triangle. Write the three equations of the planes through the vertices perpendicular to the opposite sides. Show that the third of these can be derived as a combination of the other two; and hence infer that the three planes have a line in common and that the perpendiculars from the vertices of a triangle meet in a point.

8. Solve the problem analogous to Ex. 7 for the perpendicular bisectors of the sides.

9. Note that the length of a vector is $\sqrt{\alpha\cdot\alpha}$. If α, β, and $\gamma = \beta - \alpha$ are the three sides of a triangle, expand $\gamma\cdot\gamma = (\beta - \alpha)\cdot(\beta - \alpha)$ to obtain the law of cosines.

10. Show that the sum of the squares of the diagonals of a parallelogram equals the sum of the squares of the sides. What does the difference of the squares of the diagonals equal?

11. Show that $\dfrac{\alpha\cdot\beta}{\alpha\cdot\alpha}\,\alpha$ and $\dfrac{(\alpha\times\beta)\times\alpha}{\alpha\cdot\alpha}$ are the components of β parallel and perpendicular to α by showing 1° that these vectors have the right direction, and 2° that they have the right magnitude.

12. If α, β, γ are the three edges of a parallelepiped which start from the same vertex, show that $(\alpha\times\beta)\cdot\gamma$ is the volume of the parallelepiped, the volume being considered positive if γ lies on the same side of the plane of α and β with the vector $\alpha\times\beta$.

13. Show by Ex. 12 that $(\alpha\times\beta)\cdot\gamma = \alpha\cdot(\beta\times\gamma)$ and $(\alpha\times\beta)\cdot\gamma = (\beta\times\gamma)\cdot\alpha$; and hence infer that in a product of three vectors with cross and dot, the position of the cross and dot may be interchanged and the order of the factors may be permuted cyclically without altering the value. Show that the vanishing of $(\alpha\times\beta)\cdot\gamma$ or any of its equivalent expressions denotes that α, β, γ are parallel to the same plane; the condition $\alpha\times\beta\cdot\gamma = 0$ is called the condition of complanarity.

14. Assuming $\alpha = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, $\beta = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, $\gamma = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$, expand $\alpha\cdot\gamma$, $\alpha\cdot\beta$, and $\alpha\times(\beta\times\gamma)$ in terms of the coefficients to show

$$\alpha\times(\beta\times\gamma) = (\alpha\cdot\gamma)\,\beta - (\alpha\cdot\beta)\,\gamma; \quad \text{and hence} \quad (\alpha\times\beta)\times\gamma = (\alpha\cdot\gamma)\,\beta - (\gamma\cdot\beta)\,\alpha.$$

15. The formulas of Ex. 14 for expanding a product with two crosses and the rule of Ex. 13 that a dot and a cross may be interchanged may be applied to expand

$$(\alpha\times\beta)\times(\gamma\times\delta) = (\alpha\cdot\gamma\times\delta)\,\beta - (\beta\cdot\gamma\times\delta)\,\alpha = (\alpha\times\beta\cdot\delta)\,\gamma - (\alpha\times\beta\cdot\gamma)\,\delta$$

and

$$(\alpha\times\beta)\cdot(\gamma\times\delta) = (\alpha\cdot\gamma)(\beta\cdot\delta) - (\beta\cdot\gamma)(\alpha\cdot\delta).$$

16. If α and β are two unit vectors in the xy-plane inclined at angles θ and ϕ to the x-axis, show that

$$\alpha = \mathbf{i}\cos\theta + \mathbf{j}\sin\theta, \quad \beta = \mathbf{i}\cos\phi + \mathbf{j}\sin\phi;$$

and from the fact that $\alpha\cdot\beta = \cos(\phi - \theta)$ and $\alpha\times\beta = \mathbf{k}\sin(\phi - \theta)$ obtain by multiplication the trigonometric formulas for $\sin(\phi - \theta)$ and $\cos(\phi - \theta)$.

17. If l, m, n are direction cosines, the vector $l\mathbf{i} + m\mathbf{j} + n\mathbf{k}$ is a vector of unit length in the direction for which l, m, n are direction cosines. Show that the condition for perpendicularity of two directions (l, m, n) and (l', m', n') is $ll' + mm' + nn' = 0$.

18. With the same notations as in Ex. 14 show that

$$\alpha\cdot\alpha = a_1^2 + a_2^2 + a_3^2 \quad \text{and} \quad \alpha\times\beta = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \quad \text{and} \quad \alpha\times\beta\cdot\gamma = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

19. Compute the scalar and vector products of these pairs of vectors:

$$(\alpha)\begin{cases}6\mathbf{i}+0.3\mathbf{j}-5\mathbf{k}\\0.1\mathbf{i}-4.2\mathbf{j}+2.5\mathbf{k},\end{cases}\quad(\beta)\begin{cases}\mathbf{i}+2\mathbf{j}+3\mathbf{k}\\-3\mathbf{i}-2\mathbf{j}+\mathbf{k},\end{cases}\quad(\gamma)\begin{cases}\mathbf{i}+\mathbf{k}\\\mathbf{j}+\mathbf{i}.\end{cases}$$

20. Find the areas of the parallelograms defined by the pairs of vectors in Ex. 19. Find also the sine and cosine of the angles between the vectors.

21. Prove $\alpha\times[\beta\times(\gamma\times\delta)]=(\alpha\cdot\gamma\times\delta)\beta-\alpha\cdot\beta\,\gamma\times\delta=\beta\cdot\delta\,\alpha\times\gamma-\beta\cdot\gamma\,\alpha\times\delta.$

22. What is the area of the triangle (1, 1, 1), (0, 2, 3), (0, 0, −1)?

77. Vector differentiation. As the fundamental rules of differentiation depend on the laws of subtraction, multiplication by a number, the distributive law, and the rules permitting rearrangement, it follows that the rules must be applicable to expressions containing vectors without any changes except those implied by the fact that $\alpha\times\beta\neq\beta\times\alpha$. As an illustration consider the application of the definition of differentiation to the vector product $\mathbf{u}\times\mathbf{v}$ of two vectors which are supposed to be functions of a numerical variable, say x. Then

$$\Delta(\mathbf{u}\times\mathbf{v})=(\mathbf{u}+\Delta\mathbf{u})\times(\mathbf{v}+\Delta\mathbf{v})-\mathbf{u}\times\mathbf{v}$$
$$=\mathbf{u}\times\Delta\mathbf{v}+\Delta\mathbf{u}\times\mathbf{v}+\Delta\mathbf{u}\times\Delta\mathbf{v},$$

$$\frac{\Delta(\mathbf{u}\times\mathbf{v})}{\Delta x}=\mathbf{u}\times\frac{\Delta\mathbf{v}}{\Delta x}+\frac{\Delta\mathbf{u}}{\Delta x}\times\mathbf{v}+\frac{\Delta\mathbf{u}\times\Delta\mathbf{v}}{\Delta x},$$

$$\frac{d(\mathbf{u}\times\mathbf{v})}{dx}=\lim_{\Delta x\doteq 0}\frac{\Delta(\mathbf{u}\times\mathbf{v})}{\Delta x}=\mathbf{u}\times\frac{d\mathbf{v}}{dx}+\frac{d\mathbf{u}}{dx}\times\mathbf{v}.$$

Here the ordinary rule for a product is seen to hold, except that *the order of the factors must not be interchanged.*

The interpretation of the derivative is important. Let the variable vector $\mathbf{r}$ be regarded as a function of some variable, say x, and suppose $\mathbf{r}$ is laid off from an assumed origin so that, as x varies, the terminal point of $\mathbf{r}$ describes a curve. The increment $\Delta\mathbf{r}$ of $\mathbf{r}$ corresponding to Δx is a vector quantity and in fact is the chord of the curve as indicated. *The derivative*

$$\frac{d\mathbf{r}}{dx}=\lim\frac{\Delta\mathbf{r}}{\Delta x},\quad\frac{d\mathbf{r}}{ds}=\lim\frac{\Delta\mathbf{r}}{\Delta s}=\mathbf{t}\qquad(12)$$

is therefore a vector tangent to the curve; in particular if the variable x were the arc s, the derivative would have the magnitude unity and would be a unit vector tangent to the curve.

The derivative or differential of a vector of constant length is perpendicular to the vector. This follows from the fact that the vector

then describes a circle concentric with the origin. It may also be seen analytically from the equation

$$d(\mathbf{r}\cdot\mathbf{r}) = d\mathbf{r}\cdot\mathbf{r} + \mathbf{r}\cdot d\mathbf{r} = 2\,\mathbf{r}\cdot d\mathbf{r} = d\,\text{const.} = 0. \tag{43}$$

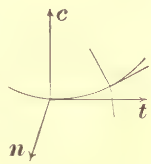
If the vector of constant length is of length unity, the increment $\Delta\mathbf{r}$ is the chord in a unit circle and, apart from infinitesimals of higher order, it is equal in magnitude to the angle subtended at the center. Consider then the derivative of the unit tangent $\mathbf{t}$ to a curve with respect to the arc s. The magnitude of $d\mathbf{t}$ is the angle the tangent turns through and the direction of $d\mathbf{t}$ is normal to $\mathbf{t}$ and hence to the curve. The vector quantity,

$$\text{curvature} \quad \mathbf{C} = \frac{d\mathbf{t}}{ds} = \frac{d^2\mathbf{r}}{ds^2}, \tag{44}$$

therefore has the magnitude of the curvature (by the definition in § 42) and the direction of the interior normal to the curve.

This work holds equally for plane or space curves. In the case of a space curve the plane which contains the tangent $\mathbf{t}$ and the curvature $\mathbf{C}$ is called the osculating plane (§ 41). By definition (§ 42) the *torsion of a space curve* is the rate of turning of the osculating plane with the arc, that is, $d\psi/ds$. To find the torsion by vector methods let $\mathbf{c}$ be a unit vector $\mathbf{C}/\sqrt{\mathbf{C}\cdot\mathbf{C}}$ along $\mathbf{C}$. Then as $\mathbf{t}$ and $\mathbf{c}$ are perpendicular, $\mathbf{n} = \mathbf{t}\times\mathbf{c}$ is a unit vector perpendicular to the osculating plane and $d\mathbf{n}$ will equal $d\psi$ in magnitude. Hence as a vector quantity the torsion is

$$\mathbf{T} = \frac{d\mathbf{n}}{ds} = \frac{d(\mathbf{t}\times\mathbf{c})}{ds} = \frac{d\mathbf{t}}{ds}\times\mathbf{c} + \mathbf{t}\times\frac{d\mathbf{c}}{ds} = \mathbf{t}\times\frac{d\mathbf{c}}{ds}, \tag{45}$$



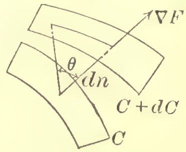
where (since $d\mathbf{t}/ds = \mathbf{C}$, and $\mathbf{c}$ is parallel to $\mathbf{C}$) the first term drops out. Next note that $d\mathbf{n}$ is perpendicular to $\mathbf{n}$ because it is the differential of a unit vector, and is perpendicular to $\mathbf{t}$ because $d\mathbf{n} = d(\mathbf{t}\times\mathbf{c}) = \mathbf{t}\times d\mathbf{c}$ and $\mathbf{t}\cdot(\mathbf{t}\times d\mathbf{c}) = 0$ since $\mathbf{t}$, $\mathbf{t}$, $d\mathbf{c}$ are necessarily complanar (Ex. 12, p. 169). Hence $\mathbf{T}$ is parallel to $\mathbf{c}$. It is convenient to consider the torsion as positive when the osculating plane seems to turn in the positive direction when viewed from the side of the normal plane upon which $\mathbf{t}$ lies. An inspection of the figure shows that in this case $d\mathbf{n}$ has the direction $-\,\mathbf{c}$ and not $+\,\mathbf{c}$. As $\mathbf{c}$ is a unit vector, the numerical value of the torsion is therefore $-\,\mathbf{c}\cdot\mathbf{T}$. Then

$$\begin{aligned} T = -\mathbf{c}\cdot\mathbf{T} &= -\,\mathbf{c}\cdot\mathbf{t}\times\frac{d\mathbf{c}}{ds} = -\,\mathbf{c}\cdot\mathbf{t}\times\frac{d}{ds}\frac{\mathbf{C}}{\sqrt{\mathbf{C}\cdot\mathbf{C}}} \\ &= -\,\mathbf{c}\cdot\mathbf{t}\times\left[\frac{d^3\mathbf{r}}{ds^3}\frac{1}{\sqrt{\mathbf{C}\cdot\mathbf{C}}} + \mathbf{C}\frac{d}{ds}\frac{1}{\sqrt{\mathbf{C}\cdot\mathbf{C}}}\right] = -\,\mathbf{c}\cdot\mathbf{t}\times\frac{d^3\mathbf{r}}{ds^3}\frac{1}{\sqrt{\mathbf{C}\cdot\mathbf{C}}} \\ &= \mathbf{t}\cdot\frac{\mathbf{C}}{\mathbf{C}\cdot\mathbf{C}}\times\frac{d^3\mathbf{r}}{ds^3} = \frac{\mathbf{r}'\cdot\mathbf{r}''\times\mathbf{r}'''}{\mathbf{r}''\cdot\mathbf{r}''}, \end{aligned} \tag{45'}$$

where differentiation with respect to s is denoted by accents.

78. Another sort of relation between vectors and differentiation comes to light in connection with the normal and directional derivatives (§ 48). If $F(x, y, z)$ is a function which has a definite value at

each point of space and if the two neighboring surfaces $F = C$ and $F = C + dC$ are considered, the normal derivative of F is the rate of change of F along the normal to the surfaces and is written dF/dn. The rate of change of F along the normal to the surface $F = C$ is more rapid than along any other direction; for the change in F between the two surfaces is $dF = dC$ and is constant, whereas the distance dn between the two surfaces is least (apart from infinitesimals of higher order) along the normal. In fact if dr denote the distance along any other direction, the relations shown by the figure are



$$dr = \sec\theta dn \quad \text{and} \quad \frac{dF}{dr} = \frac{dF}{dn}\cos\theta. \tag{46}$$

If now $\mathbf{n}$ denote a vector of unit length normal to the surface, *the product* $\mathbf{n}\,dF/dn$ *will be a vector quantity which has both the magnitude and the direction of most rapid increase of* F. Let

$$\mathbf{n}\frac{dF}{dn} = \nabla F = \text{grad } F \tag{47}$$

be the symbolic expressions for this vector, where ∇F is read as "del F" and grad F is read as "the gradient of F." If $d\mathbf{r}$ be the vector of which dr is the length, the scalar product $\mathbf{n}\cdot d\mathbf{r}$ is precisely $\cos\theta dr$, and hence it follows that

$$d\mathbf{r}\cdot\nabla F = dF \quad \text{and} \quad \mathbf{r}_1\cdot\nabla F = \frac{dF}{dr}, \tag{48}$$

where $\mathbf{r}_1$ is a unit vector in the direction $d\mathbf{r}$. The second of the equations shows that *the directional derivative in any direction is the component or projection of the gradient in that direction.*

From this fact the expression of the gradient may be found in terms of its components along the axes. For the derivatives of F along the axes are $\partial F/\partial x$, $\partial F/\partial y$, $\partial F/\partial z$, and as these are the components of ∇F along the directions $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, the result is

$$\nabla F = \text{grad } F = \mathbf{i}\frac{\partial F}{\partial x} + \mathbf{j}\frac{\partial F}{\partial y} + \mathbf{k}\frac{\partial F}{\partial z}.$$

Hence

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} \tag{49}$$

may be regarded as a symbolic vector-differentiating operator which when applied to F gives the gradient of F. The product

$$d\mathbf{r}\cdot\nabla F = \left(dx\frac{\partial}{\partial x} + dy\frac{\partial}{\partial y} + dz\frac{\partial}{\partial z}\right)F = dF \tag{50}$$

is immediately seen to give the ordinary expression for dF. From this form of grad F it does not appear that the gradient of a function is independent of the choice of axes, but from the manner of derivation of ∇F first given it does appear that grad F is a definite vector quantity independent of the choice of axes.

In the case of any given function F the gradient may be found by the application of the formula (49); but in many instances it may also be found by means of the important relation $d\mathbf{r}\cdot\nabla F = dF$ of (48). For instance to prove the formula $\nabla(FG) = F\nabla G + G\nabla F$, the relation may be applied as follows:

$$\begin{aligned} d\mathbf{r}\cdot\nabla(FG) &= d(FG) = FdG + GdF \\ &= Fd\mathbf{r}\cdot\nabla G + Gd\mathbf{r}\cdot\nabla F = d\mathbf{r}\cdot(F\nabla G + G\nabla F). \end{aligned}$$

Now as these equations hold for any direction $d\mathbf{r}$, the $d\mathbf{r}$ may be canceled by (35), p. 165, and the desired result is obtained.

The use of vector notations for treating assigned practical problems involving computation is not great, but for handling the general theory of such parts of physics as are essentially concerned with direct quantities, mechanics, hydromechanics, electromagnetic theories, etc., the actual use of the vector algorisms considerably shortens the formulas and has the added advantage of operating directly upon the magnitudes involved. At this point some of the elements of mechanics will be developed.

79. According to Newton's Second Law, when a force acts upon a particle of mass m, *the rate of change of momentum is equal to the force acting, and takes place in the direction of the force.* It therefore appears that the rate of change of momentum and momentum itself are to be regarded as vector or directed magnitudes in the application of the Second Law. Now if the vector $\mathbf{r}$, laid off from a fixed origin to the point at which the moving mass m is situated at any instant of time t, be differentiated with respect to the time t, the derivative $d\mathbf{r}/dt$ is a vector, tangent to the curve in which the particle is moving and of magnitude equal to ds/dt or v, the velocity of motion. As vectors*, then, the velocity $\mathbf{v}$ and the momentum and the force may be written as

$$\mathbf{v} = \frac{d\mathbf{r}}{dt}, \qquad m\mathbf{v}, \qquad \mathbf{F} = \frac{d}{dt}(m\mathbf{v}). \tag{51}$$

Hence $$\mathbf{F} = m\frac{d\mathbf{v}}{dt} = m\frac{d^2\mathbf{r}}{dt^2} = m\mathbf{f} \quad \text{if} \quad \mathbf{f} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2}.$$

From the equations it appears that the force $\mathbf{F}$ is the product of the mass m by a vector $\mathbf{f}$ which is the rate of change of the velocity regarded

* In applications, it is usual to denote vectors by heavy type and to denote the magnitudes of those vectors by corresponding italic letters.

as a vector. The vector $\mathbf{f}$ is called the *acceleration*: it must not be confused with the rate of change dv/dt or d^2s/dt^2 of the speed or magnitude of the velocity. The components f_x, f_y, f_z of the acceleration along the axes are the projections of $\mathbf{f}$ along the directions $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and may be written as $\mathbf{f}\cdot\mathbf{i}$, $\mathbf{f}\cdot\mathbf{j}$, $\mathbf{f}\cdot\mathbf{k}$. Then by the laws of differentiation it follows that

$$f_x = \mathbf{f}\cdot\mathbf{i} = \frac{d\mathbf{v}}{dt}\cdot\mathbf{i} = \frac{d(\mathbf{v}\cdot\mathbf{i})}{dt} = \frac{dv_x}{dt},$$

or

$$f_x = \mathbf{f}\cdot\mathbf{i} = \frac{d^2\mathbf{r}}{dt^2}\cdot\mathbf{i} = \frac{d^2(\mathbf{r}\cdot\mathbf{i})}{dt^2} = \frac{d^2x}{dt^2}.$$

Hence

$$f_x = \frac{d^2x}{dt^2}, \qquad f_y = \frac{d^2y}{dt^2}, \qquad f_z = \frac{d^2z}{dt^2},$$

and it is seen that the components of the acceleration are the accelerations of the components. If X, Y, Z are the components of the force, the equations of motion in rectangular coördinates are

$$m\frac{d^2x}{dt^2} = X, \qquad m\frac{d^2y}{dt^2} = Y, \qquad m\frac{d^2z}{dt^2} = Z. \tag{52}$$

Instead of resolving the acceleration, force, and displacement along the axes, it may be convenient to resolve them along the tangent and normal to the curve. The velocity $\mathbf{v}$ may be written as $v\mathbf{t}$, where v is the magnitude of the velocity and $\mathbf{t}$ is a unit vector tangent to the curve. Then

$$\mathbf{f} = \frac{d\mathbf{v}}{dt} = \frac{d(v\mathbf{t})}{dt} = \frac{dv}{dt}\mathbf{t} + v\frac{d\mathbf{t}}{dt}.$$

But

$$\frac{d\mathbf{t}}{dt} = \frac{d\mathbf{t}}{ds}\frac{ds}{dt} = \mathbf{C}v = \frac{v}{R}\mathbf{n}, \tag{53}$$

where R is the radius of curvature and $\mathbf{n}$ is a unit normal. Hence

$$\mathbf{f} = \frac{d^2s}{dt^2}\mathbf{t} + \frac{v^2}{R}\mathbf{n}, \qquad f_t = \frac{d^2s}{dt^2}, \qquad f_n = \frac{v^2}{R}. \tag{53'}$$

It therefore is seen that the component of the acceleration along the tangent is d^2s/dt^2, or the rate of change of the velocity regarded as a number, and the component normal to the curve is v^2/R. If T and N are the components of the force along the tangent and normal to the curve of motion, the equations are

$$T = mf_t = m\frac{d^2s}{dt^2}, \qquad N = mf_n = m\frac{v^2}{R}.$$

It is noteworthy that the force must lie in the osculating plane.

If $\mathbf{r}$ and $\mathbf{r} + \Delta\mathbf{r}$ are two positions of the radius vector, the area of the sector included by them is (except for infinitesimals of higher order)

$\Delta \mathbf{A} = \frac{1}{2}\mathbf{r}\times(\mathbf{r} + \Delta\mathbf{r}) = \frac{1}{2}\mathbf{r}\times\Delta\mathbf{r}$, and is a vector quantity of which the direction is normal to the plane of $\mathbf{r}$ and $\mathbf{r} + \Delta\mathbf{r}$, that is, to the plane through the origin tangent to the curve. The rate of description of area, or the *areal velocity*, is therefore

$$\frac{d\mathbf{A}}{dt} = \lim \frac{1}{2}\mathbf{r}\times\frac{\Delta\mathbf{r}}{\Delta t} = \frac{1}{2}\mathbf{r}\times\frac{d\mathbf{r}}{dt} = \frac{1}{2}\mathbf{r}\times\mathbf{v}. \tag{54}$$

The projections of the areal velocities on the coördinate planes, which are the same as the areal velocities of the projection of the motion on those planes, are (Ex. 11 below)

$$\frac{1}{2}\left(y\frac{dz}{dt} - z\frac{dy}{dt}\right), \qquad \frac{1}{2}\left(z\frac{dx}{dt} - x\frac{dz}{dt}\right), \qquad \frac{1}{2}\left(x\frac{dy}{dt} - y\frac{dx}{dt}\right). \tag{54'}$$

If the force $\mathbf{F}$ acting on the mass m passes through the origin, then $\mathbf{r}$ and $\mathbf{F}$ lie along the same direction and $\mathbf{r}\times\mathbf{F} = 0$. The equation of motion may then be integrated at sight.

$$m\frac{d\mathbf{v}}{dt} = \mathbf{F}, \qquad m\mathbf{r}\times\frac{d\mathbf{v}}{dt} = \mathbf{r}\times\mathbf{F} = 0,$$

$$\mathbf{r}\times\frac{d\mathbf{v}}{dt} = \frac{d}{dt}(\mathbf{r}\times\mathbf{v}) = 0, \qquad \mathbf{r}\times\mathbf{v} = \text{const.}$$

It is seen that in this case the rate of description of area is a constant vector, which means that the rate is not only constant in magnitude but is constant in direction, that is, the path of the particle m must lie in a plane through the origin. When the force passes through a fixed point, as in this case, the force is said to be *central*. Therefore when a particle moves under the action of a central force, the motion takes place in a plane passing through the center and the rate of description of areas, or the areal velocity, is constant.

80. If there are several particles, say n, in motion, each has its own equation of motion. These equations may be combined by addition and subsequent reduction.

$$m_1\frac{d^2\mathbf{r}_1}{dt^2} = \mathbf{F}_1, \quad m_2\frac{d^2\mathbf{r}_2}{dt^2} = \mathbf{F}_2, \quad \cdots, \quad m_n\frac{d^2\mathbf{r}_n}{dt^2} = \mathbf{F}_n,$$

and $$m_1\frac{d^2\mathbf{r}_1}{dt^2} + m_2\frac{d^2\mathbf{r}_2}{dt^2} + \cdots + m_n\frac{d^2\mathbf{r}_n}{dt^2} = \mathbf{F}_1 + \mathbf{F}_2 + \cdots + \mathbf{F}_n.$$

But $$m_1\frac{d^2\mathbf{r}_1}{dt^2} + m_2\frac{d^2\mathbf{r}_2}{dt^2} + \cdots + m_n\frac{d^2\mathbf{r}_n}{dt^2} = \frac{d^2}{dt^2}(m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_n\mathbf{r}_n).$$

Let $$m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_n\mathbf{r}_n = (m_1 + m_2 + \cdots + m_n)\,\bar{\mathbf{r}} = M\,\bar{\mathbf{r}}$$

or $$\bar{\mathbf{r}} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_n\mathbf{r}_n}{m_1 + m_2 + \cdots + m_n} = \frac{\Sigma m\mathbf{r}}{\Sigma m} = \frac{\Sigma m\mathbf{r}}{M}.$$

Then $$M\frac{d^2\bar{\mathbf{r}}}{dt^2} = \mathbf{F}_1 + \mathbf{F}_2 + \cdots + \mathbf{F}_n = \sum\mathbf{F}. \tag{55}$$

Now the vector $\mathbf{r}$ which has been here introduced is the vector of the center of mass or center of gravity of the particles (Ex. 5, p. 168). The result (55) states, on comparison with (51), that the center of gravity of the n masses moves as if all the mass M were concentrated at it and all the forces applied there.

The force $\mathbf{F}_i$ acting on the ith mass may be wholly or partly due to attractions, repulsions, pressures, or other actions exerted on that mass by one or more of the other masses of the system of n particles. In fact let $\mathbf{F}_i$ be written as

$$\mathbf{F}_i = \mathbf{F}_{i0} + \mathbf{F}_{i1} + \mathbf{F}_{i2} + \cdots + \mathbf{F}_{in},$$

where $\mathbf{F}_{ij}$ is the force exerted on m_i by m_j and $\mathbf{F}_{i0}$ is the force due to some agency external to the n masses which form the system. Now by Newton's Third Law, when one particle acts upon a second, the second reacts upon the first with a force which is equal in magnitude and opposite in direction. Hence to $\mathbf{F}_{ij}$ above there will correspond a force $\mathbf{F}_{ji} = - \mathbf{F}_{ij}$ exerted by m_i on m_j. In the sum $\Sigma\mathbf{F}_i$ all these equal and opposite actions and reactions will drop out and $\Sigma\mathbf{F}_i$ may be replaced by $\Sigma\mathbf{F}_{i0}$, the sum of the external forces. Hence the important theorem that: *The motion of the center of mass of a set of particles is as if all the mass were concentrated there and all the external forces were applied there* (the internal forces, that is, the forces of mutual action and reaction between the particles being entirely neglected).

The *moment of a force* about a given point is defined as the product of the force by the perpendicular distance of the force from the point. If $\mathbf{r}$ is the vector from the point as origin to any point in the line of the force, the moment is therefore $\mathbf{r}\times\mathbf{F}$ when considered as a vector quantity, and is perpendicular to the plane of the line of the force and the origin. The equations of n moving masses may now be combined in a different way and reduced. Multiply the equations by $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_n$ and add. Then

$$m_1\mathbf{r}_1\times\frac{d\mathbf{v}_1}{dt} + m_2\mathbf{r}_2\times\frac{d\mathbf{v}_2}{dt} + \cdots + m_n\mathbf{r}_n\times\frac{d\mathbf{v}_n}{dt} = \mathbf{r}_1\times\mathbf{F}_1 + \mathbf{r}_2\times\mathbf{F}_2 + \cdots + \mathbf{r}_n\times\mathbf{F}_n$$

or $$m_1\frac{d}{dt}\mathbf{r}_1\times\mathbf{v}_1 + m_2\frac{d}{dt}\mathbf{r}_2\times\mathbf{v}_2 + \cdots + m_n\frac{d}{dt}\mathbf{r}_n\times\mathbf{v}_n = \mathbf{r}_1\times\mathbf{F}_1 + \mathbf{r}_2\times\mathbf{F}_2 + \cdots + \mathbf{r}_n\times\mathbf{F}_n$$

or $$\frac{d}{dt}(m_1\mathbf{r}_1\times\mathbf{v}_1 + m_2\mathbf{r}_2\times\mathbf{v}_2 + \cdots + m_n\mathbf{r}_n\times\mathbf{v}_n) = \Sigma\mathbf{r}\times\mathbf{F}. \tag{56}$$

This equation shows that if the areal velocities of the different masses are multiplied by those masses, and all added together, the derivative of the sum obtained is equal to the moment of all the forces about the origin, the moments of the different forces being added as vector quantities.

This result may be simplified and put in a different form. Consider again the resolution of $\mathbf{F}_i$ into the sum $\mathbf{F}_{i0} + \mathbf{F}_{i1} + \cdots + \mathbf{F}_{in}$, and in particular consider the action $\mathbf{F}_{ij}$ and the reaction $\mathbf{F}_{ji} = - \mathbf{F}_{ij}$ between two particles. Let it be assumed that the action and reaction are not only equal and opposite, but lie along the line connecting the two particles. Then the perpendicular distances from the origin to the action and reaction are equal and the moments of the action and reaction are equal and opposite, and may be dropped from the sum $\Sigma\mathbf{r}_i\times\mathbf{F}_i$, which then reduces to $\Sigma\mathbf{r}_i\times\mathbf{F}_{i0}$. On the other hand a term like $m_i\mathbf{r}_i\times\mathbf{v}_i$ may be written as $\mathbf{r}_i\times(m_i\mathbf{v}_i)$. This product is formed from the momentum in exactly the same way that the moment is formed from the force, and it is called the moment of momentum. Hence the equation (56) becomes

$$\frac{d}{dt}\text{(total moment of momentum)} = \text{moment of external forces.}$$

Hence the result that, as vector quantities: *The rate of change of the moment of momentum of a system of particles is equal to the moment of the external forces* (the forces between the masses being entirely neglected under the assumption that action and reaction lie along the line connecting the masses).

EXERCISES

1. Apply the definition of differentiation to prove

(α) $d(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot d\mathbf{v} + \mathbf{v}\cdot d\mathbf{u}$, (β) $d[\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})] = d\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) + \mathbf{u}\cdot(d\mathbf{v}\times\mathbf{w}) + \mathbf{u}\cdot(\mathbf{v}\times d\mathbf{w})$.

2. Differentiate under the assumption that vectors denoted by early letters of the alphabet are constant and those designated by the later letters are variable:

(α) $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})$, (β) $\mathbf{a}\cos t + \mathbf{b}\sin t$, (γ) $(\mathbf{u}\cdot\mathbf{u})\,\mathbf{u}$,

(δ) $\mathbf{u}\times\dfrac{d\mathbf{u}}{dx}$, (ϵ) $\mathbf{u}\cdot\left(\dfrac{d\mathbf{u}}{dx}\times\dfrac{d^2\mathbf{u}}{dx^2}\right)$, (ζ) $\mathbf{c}(\mathbf{a}\cdot\mathbf{u})$.

3. Apply the rules for change of variable to show that $\dfrac{d^2\mathbf{r}}{ds^2} = \dfrac{\mathbf{r}''s' - \mathbf{r}'s''}{s'^3}$, where accents denote differentiation with respect to x. In case $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$ show that $1/\sqrt{\mathbf{C}\cdot\mathbf{C}}$ takes the usual form for the radius of curvature of a plane curve.

4. The equation of the helix is $r = \mathbf{i}a\cos\phi + \mathbf{j}a\sin\phi + \mathbf{k}b\phi$ with $s = \sqrt{a^2 + b^2}\,\phi$; show that the radius of curvature is $(a^2 + b^2)/a$.

5. Find the torsion of the helix. It is $b/(a^2 + b^2)$.

6. Change the variable from s to some other variable t in the formula for torsion.

7. In the following cases find the gradient either by applying the formula which contains the partial derivatives, or by using the relation $d\mathbf{r}\cdot\nabla F = dF$, or both:

(α) $\mathbf{r}\cdot\mathbf{r} = x^2 + y^2 + z^2$, (β) $\log r$, (γ) $r = \sqrt{\mathbf{r}\cdot\mathbf{r}}$,

(δ) $\log(x^2 + y^2) = \log[\mathbf{r}\cdot\mathbf{r} - (\mathbf{k}\cdot\mathbf{r})^2]$, (ϵ) $(\mathbf{r}\times\mathbf{a})\cdot(\mathbf{r}\times\mathbf{b})$.

8. Prove these laws of operation with the symbol ∇:

(α) $\nabla(F + G) = \nabla F + \nabla G$, (β) $G^2\nabla(F/G) = G\nabla F - F\nabla G$.

9. If r, ϕ are polar coördinates in a plane and $\mathbf{r}_1$ is a unit vector along the radius vector, show that $d\mathbf{r}_1/dt = \mathbf{n}\,d\phi/dt$ where $\mathbf{n}$ is a unit vector perpendicular to the radius. Thus differentiate $\mathbf{r} = r\mathbf{r}_1$ twice and separate the result into components along the radius vector and perpendicular to it so that

$$f_r = \frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2, \qquad f_\phi = r\frac{d^2\phi}{dt^2} + 2\frac{d\phi}{dt}\frac{dr}{dt} = \frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right).$$

10. Prove conversely to the text that if the vector rate of description of area is constant, the force must be central, that is, $\mathbf{r}\times\mathbf{F} = 0$.

11. Note that $\mathbf{r}\times\mathbf{v}\cdot\mathbf{i}$, $\mathbf{r}\times\mathbf{v}\cdot\mathbf{j}$, $\mathbf{r}\times\mathbf{v}\cdot\mathbf{k}$ are the projections of the areal velocities upon the planes $x = 0$, $y = 0$, $z = 0$. Hence derive (54′) of the text.

12. Show that the Cartesian expressions for the magnitude of the velocity and of the acceleration and for the rate of change of the speed dv/dt are

$$v = \sqrt{x'^2 + y'^2 + z'^2}, \quad f = \sqrt{x''^2 + y''^2 + z''^2}, \quad v' = \frac{x'x'' + y'y'' + z'z''}{\sqrt{x'^2 + y'^2 + z'^2}},$$

where accents denote differentiation with respect to the time.

13. Suppose that a body which is rigid is rotating about an axis with the angular velocity $\omega = d\phi/dt$. Represent the angular velocity by a vector $\mathbf{a}$ drawn along the axis and of magnitude equal to ω. Show that the velocity of any point in space is $\mathbf{v} = \mathbf{a}\times\mathbf{r}$, where $\mathbf{r}$ is the vector drawn to that point from any point of the axis as origin. Show that the acceleration of the point determined by $\mathbf{r}$ is in a plane through the point and perpendicular to the axis, and that the components are

$$\mathbf{a}\times(\mathbf{a}\times\mathbf{r}) = (\mathbf{a}\cdot\mathbf{r})\,\mathbf{a} - \omega^2\mathbf{r} \text{ toward the axis,} \quad (d\mathbf{a}/dt)\times\mathbf{r} \text{ perpendicular to the axis,}$$

under the assumption that the axis of rotation is invariable.

14. Let $\bar{\mathbf{r}}$ denote the center of gravity of a system of particles and $\mathbf{r}_i'$ denote the vector drawn from the center of gravity to the ith particle so that $\mathbf{r}_i = \bar{\mathbf{r}} + \mathbf{r}_i'$ and $\mathbf{v}_i = \bar{\mathbf{v}} + \mathbf{v}_i'$. The kinetic energy of the ith particle is by definition

$$\tfrac{1}{2} m_i v_i^2 = \tfrac{1}{2} m_i \mathbf{v}_i\cdot\mathbf{v}_i = \tfrac{1}{2} m_i (\bar{\mathbf{v}} + \mathbf{v}_i')\cdot(\bar{\mathbf{v}} + \mathbf{v}_i').$$

Sum up for all particles and simplify by using the fact $\Sigma m_i\mathbf{r}_i' = 0$, which is due to the assumption that the origin for the vectors $\mathbf{r}_i'$ is at the center of gravity. Hence prove the important theorem: *The total kinetic energy of a system is equal to the kinetic energy which the total mass would have if moving with the center of gravity plus the energy computed from the motion relative to the center of gravity as origin*, that is,

$$T = \tfrac{1}{2}\Sigma m_i v_i^2 = \tfrac{1}{2} M\bar{v}^2 + \tfrac{1}{2}\Sigma m_i v_i'^2.$$

15. Consider a rigid body moving in a plane, which may be taken as the xy-plane. Let any point $\mathbf{r}_0$ of the body be marked and other points be denoted relative to it by $\mathbf{r}'$. The motion of any point $\mathbf{r}'$ is compounded from the motion of $\mathbf{r}_0$ and from the angular velocity $\mathbf{a} = \mathbf{k}\omega$ of the body about the point $\mathbf{r}_0$. In fact the velocity $\mathbf{v}$ of any point is $\mathbf{v} = \mathbf{v}_0 + \mathbf{a}\times\mathbf{r}'$. Show that the velocity of the point denoted by $\mathbf{r}' = \mathbf{k}\times\mathbf{v}_0/\omega$ is zero. This point is known as the instantaneous center of rotation (§ 39). Show that the coördinates of the instantaneous center referred to axes at the origin of the vectors $\mathbf{r}$ are

$$x = \mathbf{r}\cdot\mathbf{i} = x_0 - \frac{1}{\omega}\frac{dy_0}{dt}, \qquad y = \mathbf{r}\cdot\mathbf{j} = y_0 + \frac{1}{\omega}\frac{dx_0}{dt}.$$

16. If several forces $\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_n$ act on a body, the sum $\mathbf{R} = \Sigma\mathbf{F}_i$ is called the *resultant* and the sum $\Sigma\mathbf{r}_i\times\mathbf{F}_i$, where $\mathbf{r}_i$ is drawn from an origin O to a point in the line of the force $\mathbf{F}_i$, is called the *resultant moment* about O. Show that the resultant moments $\mathbf{M}_O$ and $\mathbf{M}_{O'}$ about two points are connected by the relation $\mathbf{M}_{O'} = \mathbf{M}_O + \mathbf{M}_{O'}(\mathbf{R}_O)$, where $\mathbf{M}_{O'}(\mathbf{R}_O)$ means the moment about O' of the resultant $\mathbf{R}$ considered as applied at O. Infer that moments about all points of any line parallel to the resultant are equal. Show that in any plane perpendicular to $\mathbf{R}$ there is a point O' given by $\mathbf{r} = \mathbf{R}\times\mathbf{M}_O/\mathbf{R}\cdot\mathbf{R}$, where O is any point of the plane, such that $\mathbf{M}_{O'}$ is parallel to $\mathbf{R}$.

PART II. DIFFERENTIAL EQUATIONS

CHAPTER VII

GENERAL INTRODUCTION TO DIFFERENTIAL EQUATIONS

81. Some geometric problems. The application of the differential calculus to plane curves has given a means of determining some geometric properties of the curves. For instance, the length of the subnormal of a curve (§ 7) is $y\,dy/dx$, which in the case of the parabola $y^2 = 4\,px$ is $2\,p$, that is, the subnormal is constant. Suppose now it were desired conversely to find all curves for which the subnormal is a given constant m. The statement of this problem is evidently contained in the equation

$$y\frac{dy}{dx} = m \quad \text{or} \quad yy' = m \quad \text{or} \quad y\,dy = m\,dx.$$

Again, the radius of curvature of the lemniscate $r^2 = a^2 \cos 2\,\phi$ is found to be $R = a^2/3\,r$, that is, the radius of curvature varies inversely as the radius. If conversely it were desired to find all curves for which the radius of curvature varies inversely as the radius of the curve, the statement of the problem would be the equation

$$\frac{\left[r^2 + \left(\dfrac{dr}{d\phi}\right)^2\right]^{\frac{3}{2}}}{r^2 - r\dfrac{d^2r}{d\phi^2} + 2\left(\dfrac{dr}{d\phi}\right)^2} = \frac{k}{r},$$

where k is a constant called a factor of proportionality.*

Equations like these are unlike ordinary algebraic equations because, in addition to the variables x, y or r, ϕ and certain constants m or k, they contain also derivatives, as dy/dx or $dr/d\phi$ and $d^2r/d\phi^2$, of one of the variables with respect to the other. An equation which contains

* Many problems in geometry, mechanics, and physics are stated in terms of variation. For purposes of analysis the statement x varies as y, or $x \propto y$, is written as $x = ky$, introducing a constant k called a factor of proportionality to convert the variation into an equation. In like manner the statement x varies inversely as y, or $x \propto 1/y$, becomes $x = k/y$, and x varies jointly with y and z becomes $x = kyz$.

derivatives is called a *differential equation*. The *order* of the differential equation is the order of the highest derivative it contains. The equations above are respectively of the first and second orders. A differential equation of the first order may be symbolized as $\Phi(x, y, y') = 0$, and one of the second order as $\Phi(x, y, y', y'') = 0$. A function $y = f(x)$ given explicitly or defined implicitly by the relation $F(x, y) = 0$ is said to be a *solution* of a given differential equation if the equation is true for all values of the independent variable x when the expressions for y and its derivatives are substituted in the equation.

Thus to show that (no matter what the value of a is) the relation

$$4\,ay - x^2 + 2\,a^2 \log x = 0$$

gives a solution of the differential equation of the second order

$$1 + \left(\frac{dy}{dx}\right)^2 - x^2\left(\frac{d^2y}{dx^2}\right)^2 = 0,$$

it is merely necessary to form the derivatives

$$2\,a\frac{dy}{dx} = x - \frac{a^2}{x}, \qquad 2\,a\frac{d^2y}{dx^2} = 1 + \frac{a^2}{x^2}$$

and substitute them in the given equation together with y to see that

$$1 + \left(\frac{dy}{dx}\right)^2 - x^2\left(\frac{d^2y}{dx^2}\right)^2 = 1 + \frac{1}{4\,a^2}\left(x^2 - 2\,a^2 + \frac{a^4}{x^2}\right) - \frac{x^2}{4\,a^2}\left(1 + \frac{2\,a^2}{x^2} + \frac{a^4}{x^4}\right) = 0$$

is clearly satisfied for all values of x. It appears therefore that the given relation for y is a solution of the given equation.

To *integrate* or *solve* a differential equation is to find all the functions which satisfy the equation. Geometrically speaking, it is to find all the curves which have the property expressed by the equation. In mechanics it is to find all possible motions arising from the given forces. The method of integrating or solving a differential equation depends largely upon the *ingenuity* of the solver. In many cases, however, some method is immediately obvious. For instance if it be possible to *separate the variables*, so that the differential dy is multiplied by a function of y alone and dx by a function of x alone, as in the equation

$$\phi(y)\,dy = \psi(x)\,dx, \quad \text{then} \quad \int \phi(y)\,dy = \int \psi(x)\,dx + C \qquad (1)$$

will clearly be the integral or solution of the differential equation.

As an example, let the curves of constant subnormal be determined. Here

$$ydy = mdx \quad \text{and} \quad y^2 = 2\,mx + C.$$

The variables are already separated and the integration is immediate. The curves are parabolas with semi-latus rectum equal to the constant and with the axis

coincident with the axis of x. If in particular it were desired to determine that curve whose subnormal was m and which passed through the origin, it would merely be necessary to substitute $(0, 0)$ in the equation $y^2 = 2\,mx + C$ to ascertain what particular value must be assigned to C in order that the curve pass through $(0, 0)$. The value is $C = 0$.

Another example might be to determine the curves for which the x-intercept varies as the abscissa of the point of tangency. As the expression (§ 7) for the x-intercept is $x - y\,dx/dy$, the statement is

$$x - y\frac{dx}{dy} = kx \quad \text{or} \quad (1-k)\,x = y\frac{dx}{dy}.$$

Hence $$(1-k)\frac{dy}{y} = \frac{dx}{x} \quad \text{and} \quad (1-k)\log y = \log x + C.$$

If desired, this expression may be changed to another form by using each side of the equality as an exponent with the base e. Then

$$e^{(1-k)\log y} = e^{\log x + C} \quad \text{or} \quad y^{1-k} = e^C x = C'x.$$

As C is an arbitrary constant, the constant $C' = e^C$ is also arbitrary and the solution may simply be written as $y^{1-k} = Cx$, where the accent has been omitted from the constant. If it were desired to pick out that particular curve which passed through the point $(1, 1)$, it would merely be necessary to determine C from the equation

$$1^{1-k} = C\,1, \quad \text{and hence} \quad C = 1.$$

As a third example let the curves whose tangent is constant and equal to a be determined. The length of the tangent is $y\sqrt{1 + y'^2}/y'$ and hence the equation is

$$y\frac{\sqrt{1 + y'^2}}{y'} = a \quad \text{or} \quad y^2\frac{1 + y'^2}{y'^2} = a^2 \quad \text{or} \quad 1 = -\frac{\sqrt{a^2 - y^2}}{y}\,y'.$$

The variables are therefore separable and the results are

$$dx = -\frac{\sqrt{a^2 - y^2}}{y}\,dy \quad \text{and} \quad x + C = \sqrt{a^2 - y^2} - a\log\frac{a + \sqrt{a^2 - y^2}}{y}.$$

If it be desired that the tangent at the origin be vertical so that the curve passes through $(0, a)$, the constant C is 0. The curve is the tractrix or "curve of pursuit" as described by a calf dragged at the end of a rope by a person walking along a straight line.

82. Problems which involve the radius of curvature will lead to differential equations of the second order. The method of solving such problems is to *reduce the equation, if possible, to one of the first order.* For the second derivative may be written as

$$y'' = \frac{dy'}{dx} = \frac{dy'}{dy}\,y', \tag{2}$$

and $$R = \frac{(1 + y'^2)^{\frac{3}{2}}}{y''} = \frac{(1 + y'^2)^{\frac{3}{2}}}{\dfrac{dy'}{dx}} = \frac{(1 + y'^2)^{\frac{3}{2}}}{y'\dfrac{dy'}{dy}}. \tag{2'}$$

is the expression for the radius of curvature. If it be given that the radius of curvature is of the form $f(x)\phi(y')$ or $f(y)\phi(y')$,

$$\frac{(1+y'^2)^{\frac{3}{2}}}{\frac{dy'}{dx}} = f(x)\phi(y') \quad \text{or} \quad \frac{(1+y'^2)^{\frac{3}{2}}}{y'\frac{dy'}{dy}} = f(y)\phi(y'), \qquad (3)$$

the variables x and y' or y and y' are immediately separable, and an integration may be performed. This will lead to an equation of the first order; and if the variables are again separable, the solution may be completed by the methods of the above examples.

In the first place consider curves whose radius of curvature is constant. Then

$$\frac{(1+y'^2)^{\frac{3}{2}}}{\frac{dy'}{dx}} = a \quad \text{or} \quad \frac{dy'}{(1+y'^2)^{\frac{3}{2}}} = \frac{dx}{a} \quad \text{and} \quad \frac{y'}{\sqrt{1+y'^2}} = \frac{x-C}{a},$$

where the constant of integration has been written as $-C/a$ for future convenience. The equation may now be solved for y' and the variables become separated with the results

$$y' = \frac{x-C}{\sqrt{a^2-(x-C)^2}} \quad \text{or} \quad dy = \frac{(x-C)}{\sqrt{a^2-(x-C)^2}}dx.$$

Hence $\qquad y - C' = -\sqrt{a^2-(x-C)^2} \quad \text{or} \quad (x-C)^2 + (y-C')^2 = a^2.$

The curves, as should be anticipated, are circles of radius a and with any arbitrary point (C, C') as center. It should be noted that, as the solution has required two successive integrations, there are two arbitrary constants C and C' of integration in the result.

As a second example consider the curves whose radius of curvature is double the normal. As the length of the normal is $y\sqrt{1+y'^2}$, the equation becomes

$$\frac{(1+y'^2)^{\frac{3}{2}}}{y'\frac{dy'}{dy}} = 2y\sqrt{1+y'^2} \quad \text{or} \quad \frac{1+y'^2}{y'\frac{dy'}{dy}} = \pm 2y,$$

where the double sign has been introduced when the radical is removed by cancellation. This is necessary; for before the cancellation the signs were ambiguous and there is no reason to assume that the ambiguity disappears. In fact, if the curve is concave up, the second derivative is positive and the radius of curvature is reckoned as positive, whereas the normal is positive or negative according as the curve is above or below the axis of x; similarly, if the curve is concave down. Let the negative sign be chosen. This corresponds to a curve above the axis and concave down or below the axis and concave up, that is, the normal and the radius of curvature have the same direction. Then

$$\frac{dy}{y} = -\frac{2y'dy'}{1+y'^2} \quad \text{and} \quad \log y = -\log(1+y'^2) + \log 2C,$$

where the constant has been given the form $\log 2C$ for convenience. This expression may be thrown into algebraic form by exponentiation, solved for y', and then

$$y(1+y'^2) = 2C \quad \text{or} \quad y'^2 = \frac{2C - y}{y} \quad \text{or} \quad \frac{y\,dy}{\sqrt{2Cy - y^2}} = dx.$$

Hence
$$x - C' = C \operatorname{vers}^{-1}\frac{y}{C} - \sqrt{2Cy - y^2}.$$

The curves are cycloids of which the generating circle has an arbitrary radius C and of which the cusps are upon the x-axis at the points $C' \pm 2k\pi C$. If the positive sign had been taken in the equation, the curves would have been entirely different; see Ex. 5 (α).

The number of arbitrary constants of integration which enter into the solution of a differential equation depends on the number of integrations which are performed and is equal to the order of the equation. This results in giving a family of curves, dependent on one or more parameters, as the solution of the equation. To pick out any particular member of the family, additional conditions must be given. Thus, if there is only one constant of integration, the curve may be required to pass through a given point; if there are two constants, the curve may be required to pass through a given point and have a given slope at that point, or to pass through two given points. These additional conditions are called *initial conditions*. In mechanics the initial conditions are very important; for the point reached by a particle describing a curve under the action of assigned forces depends not only on the forces, but on the point at which the particle started and the velocity with which it started. In all cases the distinction between the *constants of integration* and the *given constants of the problem* (in the foregoing examples, the distinction between C, C' and m, k, a) should be kept clearly in mind

EXERCISES

1. Verify the solutions of the differential equations:

(α) $xy + \frac{1}{2}x^2 = C$, $y + x + xy' = 0$. (β) $x^3y^3(3e^x + C) = 1$, $xy' + y + x^4y^4e^x = 0$.

(γ) $(1+x^2)y'^2 = 1$, $2x = Ce^y - C^{-1}e^{-y}$. (δ) $y + xy' = x^4y'^2$, $xy = C^2x + C$.

(ϵ) $y'' + y'/x = 0$, $y = C\log x + C_1$. (ζ) $y = Ce^x + C_1e^{2x}$, $y'' + 2y = 3y'$.

(η) $y''' - y = x^2$, $y = Ce^x + e^{-\frac{1}{2}x}\left(C_1 \cos\frac{x\sqrt{3}}{2} + C_2 \sin\frac{x\sqrt{3}}{2}\right) - x^2$.

2. Determine the curves which have the following properties:

(α) The subtangent is constant; $y^m = Ce^x$. If through (2, 2), $y^m = 2^m e^{x-2}$.

(β) The right triangle formed by the tangent, subtangent, and ordinate has the constant area $k/2$; the hyperbolas $xy + Cy + k = 0$. Show that if the curve passes through (1, 2) and (2, 1), the arbitrary constant C is 0 and the given k is -2.

(γ) The normal is constant in length; the circles $(x - C)^2 + y^2 = k^2$.

(δ) The normal varies as the square of the ordinate; catenaries $ky = \cosh k(x - C)$. If in particular the curve is perpendicular to the y-axis, $C = 0$.

(ϵ) The area of the right triangle formed by the tangent, normal, and x-axis is inversely proportional to the slope; the circles $(x - C)^2 + y^2 = k$.

3. Determine the curves which have the following properties:

(α) The angle between the radius vector and tangent is constant; spirals $r = Ce^{k\phi}$.

(β) The angle between the radius vector and tangent is half that between the radius and initial line; cardioids $r = C(1 - \cos\phi)$.

(γ) The perpendicular from the pole to a tangent is constant; $r\cos(\phi - C) = k$.

(δ) The tangent is equally inclined to the radius vector and to the initial line; the two sets of parabolas $r = C/(1 \pm \cos\phi)$.

(ϵ) The radius is equally inclined to the normal and to the initial line; circles $r = C\cos\phi$ or lines $r\cos\phi = C$.

4. The arc s of a curve is proportional to the area A, where in rectangular coördinates A is the area under the curve and in polar coördinates it is the area included by the curve and the radius vectors. From the equation $ds = dA$ show that the curves which satisfy the condition are catenaries for rectangular coördinates and lines for polar coördinates.

5. Determine the curves for which the radius of curvature

(α) is twice the normal and oppositely directed; parabolas $(x - C)^2 = C'(2y - C')$.

(β) is equal to the normal and in same direction; circles $(x - C)^2 + y^2 = C'^2$.

(γ) is equal to the normal and in opposite direction; catenaries.

(δ) varies as the cube of the normal; conics $kCy^2 - C^2(x + C')^2 = k$.

(ϵ) projected on the x-axis equals the abscissa; catenaries.

(ζ) projected on the x-axis is the negative of the abscissa; circles.

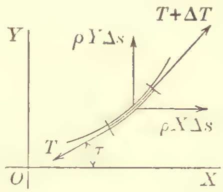
(η) projected on the x-axis is twice the abscissa.

(θ) is proportional to the slope of the tangent or of the normal.

83. Problems in mechanics and physics. In many physical problems the statement involves an equation between the *rate of change* of some quantity and the value of that quantity. In this way the solution of the problem is made to depend on the integration of a differential equation of the first order. If x denotes any quantity, the rate of increase in x is dx/dt and the rate of decrease in x is $-dx/dt$; and consequently when the rate of change of x is a function of x, the variables are immediately separated and the integration may be performed. The constant of integration has to be determined from the initial conditions; the constants inherent in the problem may be given in advance or their values may be determined by comparing x and t at some subsequent time. The exercises offered below will exemplify the treatment of such problems.

In other physical problems the statement of the question as a differential equation is not so direct and is carried out by an examination of the problem with a view to stating a relation between the increments or differentials of the dependent and independent variables, as in some geometric relations already discussed (§ 40), and in the problem of the tension in a rope wrapped around a cylindrical post discussed below.

The method may be further illustrated by the derivation of the differential equations of the curve of equilibrium of a flexible string or chain. Let ρ be the density of the chain so that $\rho\Delta s$ is the mass of the length Δs; let X and Y be the components of the force (estimated per unit mass) acting on the elements of the chain. Let T denote the tension in the chain, and τ the inclination of the element of chain. From the figure it then appears that the components of all the forces acting on Δs are



$$(T + \Delta T)\cos(\tau + \Delta\tau) - T\cos\tau + X\rho\Delta s = 0,$$
$$(T + \Delta T)\sin(\tau + \Delta\tau) - T\sin\tau + Y\rho\Delta s = 0;$$

for these must be zero if the element is to be in a position of equilibrium. The equations may be written in the form

$$\Delta(T\cos\tau) + X\rho\Delta s = 0, \qquad \Delta(T\sin\tau) + Y\rho\Delta s = 0;$$

and if they now be divided by Δs and if Δs be allowed to approach zero, the result is the two equations of equilibrium

$$\frac{d}{ds}\left(T\frac{dx}{ds}\right) + \rho X = 0, \qquad \frac{d}{ds}\left(T\frac{dy}{ds}\right) + \rho Y = 0, \tag{4}$$

where $\cos\tau$ and $\sin\tau$ are replaced by their values dx/ds and dy/ds.

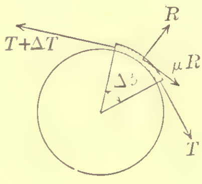
If the string is acted on only by forces parallel to a given direction, let the y-axis be taken as parallel to that direction. Then the component X will be zero and the first equation may be integrated. The result is

$$\frac{d}{ds}\left(T\frac{dx}{ds}\right) = 0, \qquad T\frac{dx}{ds} = C, \qquad T = C\frac{ds}{dx}.$$

This value of T may be substituted in the second equation. There is thus obtained a differential equation of the second order

$$\frac{d}{ds}\left(C\frac{dy}{dx}\right) + \rho Y = 0 \quad\text{or}\quad C\frac{y''}{\sqrt{1+y'^2}} + \rho Y = 0. \tag{4'}$$

If this equation can be integrated, the form of the curve of equilibrium may be found.



Another problem of a different nature in strings is to consider the variation of the tension in a rope wound around a cylinder without overlapping. The forces acting on the element Δs of the rope are the tensions T and $T + \Delta T$, the normal pressure or reaction R of the cylinder, and the force of friction which is proportional to the pressure. It will be assumed that the normal reaction lies in the angle $\Delta\phi$ and that the coefficient of friction is μ so that the force of friction is μR. The components along the radius and along the tangent are

$$(T + \Delta T)\sin \Delta\phi - R\cos(\theta\Delta\phi) - \mu R \sin(\theta\Delta\phi) = 0, \qquad 0 < \theta < 1,$$
$$(T + \Delta T)\cos\Delta\phi + R\sin(\theta\Delta\phi) - \mu R\cos(\theta\Delta\phi) - T = 0.$$

Now discard all infinitesimals except those of the first order. It must be borne in mind that the pressure R is the reaction on the infinitesimal arc Δs and hence is itself infinitesimal. The substitutions are therefore $Td\phi$ for $(T + \Delta T)\sin\Delta\phi$, R for $R\cos\theta\Delta\phi$, 0 for $R\sin\theta\Delta\phi$, and $T + dT$ for $(T + \Delta T)\cos\Delta\phi$. The equations therefore reduce to two simple equations

$$Td\phi - R = 0, \qquad dT - \mu R = 0,$$

from which the unknown R may be eliminated with the result

$$dT = \mu T d\phi \quad \text{or} \quad T = Ce^{\mu\phi} \quad \text{or} \quad T = T_0 e^{\mu\phi},$$

where T_0 is the tension when ϕ is 0. The tension therefore runs up exponentially and affords ample explanation of why a man, by winding a rope about a post, can readily hold a ship or other object exerting a great force at the other end of the rope. If μ is 1/3, three turns about the post will hold a force 535 T_0, or over 25 tons, if the man exerts a force of a hundredweight.

84. If a constant mass m is moving along a line under the influence of a force F acting along the line, Newton's Second Law of Motion (p. 13) states the problem of the motion as the differential equation

$$mf = F \quad \text{or} \quad m\frac{d^2x}{dt^2} = F \tag{5}$$

of the second order; and it therefore appears that the complete solution of a problem in rectilinear motion requires the integration of this equation. The acceleration may be written as

$$f = \frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt} = v\frac{dv}{dx};$$

and hence the equation of motion takes either of the forms

$$m\frac{dv}{dt} = F \quad \text{or} \quad mv\frac{dv}{dx} = F. \tag{5'}$$

It now appears that there are several cases in which the first integration may be performed. For if the force is a function of the velocity or of the time or a product of two such functions, the variables are separated in the first form of the equation; whereas if the force is a function of the velocity or of the coördinate x or a product of two such functions, the variables are separated in the second form of the equation.

When the first integration is performed according to either of these methods, there will arise an equation between the velocity and either the time t or the coördinate x. In this equation will be contained a constant of integration which may be determined by the initial conditions, that is, by the knowledge of the velocity at the start, whether in

time or in position. Finally it will be possible (at least theoretically) to solve the equation and express the velocity as a function of the time t or of the position x, as the case may be, and integrate a second time. The carrying through in practice of this sketch of the work will be exemplified in the following two examples.

Suppose a particle of mass m is projected vertically upward with the velocity V. Solve the problem of the motion under the assumption that the resistance of the air varies as the velocity of the particle. Let the distance be measured vertically upward. The forces acting on the particle are two, — the force of gravity which is the weight $W = mg$, and the resistance of the air which is kv. Both these forces are negative because they are directed toward diminishing values of x. Hence

$$mf = -mg - kv \quad \text{or} \quad m\frac{dv}{dt} = -mg - kv,$$

where the first form of the equation of motion has been chosen, although in this case the second form would be equally available. Then integrate.

$$\frac{dv}{g + \frac{k}{m}v} = -dt \quad \text{and} \quad \log\left(g + \frac{k}{m}v\right) = -\frac{k}{m}t + C.$$

As by the initial conditions $v = V$ when $t = 0$, the constant C is found from

$$\log\left(g + \frac{k}{m}V\right) = -\frac{k}{m}0 + C; \quad \text{hence} \quad \frac{g + \frac{k}{m}v}{g + \frac{k}{m}V} = e^{-\frac{k}{m}t}$$

is the relation between v and t found by substituting the value of C. The solution for v gives

$$v = \frac{dx}{dt} = \left(\frac{m}{k}g + V\right)e^{-\frac{k}{m}t} - \frac{m}{k}g.$$

Hence $$x = -\frac{m}{k}\left(\frac{m}{k}g + V\right)e^{-\frac{k}{m}t} - \frac{m}{k}gt + C.$$

If the particle starts from the origin $x = 0$, the constant C is found to be

$$C = \frac{m}{k}\left(\frac{m}{k}g + V\right) \quad \text{and} \quad x = \frac{m}{k}\left(\frac{m}{k}g + V\right)\left(1 - e^{-\frac{k}{m}t}\right) - \frac{m}{k}gt.$$

Hence the position of the particle is expressed in terms of the time and the problem is solved. If it be desired to find the time which elapses before the particle comes to rest and starts to drop back, it is merely necessary to substitute $v = 0$ in the relation connecting the velocity and the time, and solve for the time $t = T$; and if this value of t be substituted in the expression for x, the total distance X covered in the ascent will be found. The results are

$$T = \frac{m}{k}\log\left(1 + \frac{k}{mg}V\right), \qquad X = \left(\frac{m}{k}\right)^2\left[\frac{k}{m}V - g\log\left(1 + \frac{k}{mg}V\right)\right].$$

As a second example consider the motion of a particle vibrating up and down at the end of an elastic string held in the field of gravity. By Hooke's Law for

elastic strings the force exerted by the string is proportional to the extension of the string over its natural length, that is, $F = k\Delta l$. Let l be the length of the string, $\Delta_0 l$ the extension of the string just sufficient to hold the weight $W = mg$ at rest so that $k\Delta_0 l = mg$, and let x measured downward be the additional extension of the string at any instant of the motion. The force of gravity mg is positive and the force of elasticity $-k(\Delta_0 l + x)$ is negative. The second form of the equation of motion is to be chosen. Hence

$$mv\frac{dv}{dx} = mg - k(\Delta_0 l + x) \quad \text{or} \quad mv\frac{dv}{dx} = -kx, \quad \text{since} \quad mg = k\Delta_0 l.$$

Then $$mvdv = -kxdx \quad \text{or} \quad mv^2 = -kx^2 + C.$$

Suppose that $x = a$ is the amplitude of the motion, so that when $x = a$ the velocity $v = 0$ and the particle stops and starts back. Then $C = ka^2$. Hence

$$v = \frac{dx}{dt} = \sqrt{\frac{k}{m}}\sqrt{a^2 - x^2} \quad \text{or} \quad \frac{dx}{\sqrt{a^2 - x^2}} = \sqrt{\frac{k}{m}}\,dt,$$

and $$\sin^{-1}\frac{x}{a} = \sqrt{\frac{k}{m}}\,t + C \quad \text{or} \quad x = a\sin\left(\sqrt{\frac{k}{m}}\,t + C\right).$$

Now let the time be measured from the instant when the particle passes through the position $x = 0$. Then C satisfies the equation $0 = a\sin C$ and may be taken as zero. The motion is therefore given by the equation $x = a\sin\sqrt{k/m}\,t$ and is periodic. While t changes by $2\pi\sqrt{m/k}$ the particle completes an entire oscillation. The time $T = 2\pi\sqrt{m/k}$ is called the *periodic time*. The motion considered in this example is characterized by the fact that the total force $-kx$ is proportional to the displacement from a certain origin and is directed toward the origin. Motion of this sort is called *simple harmonic motion* (briefly S. H. M.) and is of great importance in mechanics and physics.

EXERCISES

1. The sum of \$100 is put at interest at 4 per cent per annum under the condition that the interest shall be compounded at each instant. Show that the sum will amount to \$200 in 17 yr. 4 mo., and to \$1000 in $57\frac{1}{2}$ yr.

2. Given that the rate of decomposition of an amount x of a given substance is proportional to the amount of the substance remaining undecomposed. Solve the problem of the decomposition and determine the constant of integration and the physical constant of proportionality if $x = 5.11$ when $t = 0$ and $x = 1.48$ when $t = 40$ min. *Ans.* $k = .0309$.

3. A substance is undergoing transformation into another at a rate which is assumed to be proportional to the amount of the substance still remaining untransformed. If that amount is 35.6 when $t = 1$ hr. and 13.8 when $t = 4$ hr., determine the amount at the start when $t = 0$ and the constant of proportionality and find how many hours will elapse before only one-thousandth of the original amount will remain.

4. If the activity A of a radioactive deposit is proportional to its rate of diminution and is found to decrease to $\frac{1}{2}$ its initial value in 4 days, show that A satisfies the equation $A/A_0 = e^{-.173t}$.

5. Suppose that amounts a and b respectively of two substances are involved in a reaction in which the velocity of transformation dx/dt is proportional to the product $(a-x)(b-x)$ of the amounts remaining untransformed. Integrate on the supposition that $a \neq b$.

$$\log \frac{b(a-x)}{a(b-x)} = (a-b)kt\,; \quad \text{and if} \quad \begin{array}{ccc} \underline{t} & \underline{a-x} & \underline{b-x} \\ 393 & 0.4866 & 0.2342 \\ 1265 & 0.3879 & 0.1354 \end{array}$$

determine the product $k(a-b)$.

6. Integrate the equation of Ex. 5 if $a = b$, and determine a and k if $x = 9.87$ when $t = 15$ and $x = 13.69$ when $t = 55$.

7. If the velocity of a chemical reaction in which three substances are involved is proportional to the continued product of the amounts of the substances remaining, show that the equation between x and the time is

$$\frac{\log\left(\dfrac{a}{a-x}\right)^{b-c}\left(\dfrac{b}{b-x}\right)^{c-a}\left(\dfrac{c}{c-x}\right)^{a-b}}{(a-b)(b-c)(c-a)} = -kt, \quad \text{where} \quad \begin{cases} x = 0 \\ t = 0. \end{cases}$$

8. Solve Ex. 7 if $a = b \neq c$; also when $a = b = c$. Note the very different forms of the solution in the three cases.

9. The rate at which water runs out of a tank through a small pipe issuing horizontally near the bottom of the tank is proportional to the square root of the height of the surface of the water above the pipe. If the tank is cylindrical and half empties in 30 min., show that it will completely empty in about 100 min.

10. Discuss Ex. 9 in case the tank were a right cone or frustum of a cone.

11. Consider a vertical column of air and assume that the pressure at any level is due to the weight of the air above. Show that $p = p_0 e^{-kh}$ gives the pressure at any height h, if Boyle's Law that the density of a gas varies as the pressure be used.

12. Work Ex. 11 under the assumption that the adiabatic law $p \propto \rho^{1.4}$ represents the conditions in the atmosphere. Show that in this case the pressure would become zero at a finite height. (If the proper numerical data are inserted, the height turns out to be about 20 miles. The adiabatic law seems to correspond better to the facts than Boyle's Law.)

13. Let l be the natural length of an elastic string, let Δl be the extension, and assume Hooke's Law that the force is proportional to the extension in the form $\Delta l = klF$. Let the string be held in a vertical position so as to elongate under its own weight W. Show that the elongation is $\frac{1}{2}kWl$.

14. The density of water under a pressure of p atmospheres is $\rho = 1 + 0.00004\,p$. Show that the surface of an ocean six miles deep is about 600 ft. below the position it would have if water were incompressible.

15. Show that the equations of the curve of equilibrium of a string or chain are

$$\frac{d}{ds}\left(T\frac{dr}{ds}\right) + \rho R = 0, \qquad \frac{d}{ds}\left(T\frac{rd\phi}{ds}\right) + \rho\Phi = 0$$

in polar coördinates, where R and Φ are the components of the force along the radius vector and perpendicular to it.

16. Show that $dT + \rho S ds = 0$ and $T + \rho RN = 0$ are the equations of equilibrium of a string if R is the radius of curvature and S and N are the tangential and normal components of the forces.

17.* Show that when a uniform chain is supported at two points and hangs down between the points under its own weight, the curve of equilibrium is the catenary.

18. Suppose the mass dm of the element ds of a chain is proportional to the projection dx of ds on the x-axis, and that the chain hangs in the field of gravity. Show that the curve is a parabola. (This is essentially the problem of the shape of the cables in a suspension bridge when the roadbed is of uniform linear density; for the weight of the cables is negligible compared to that of the roadbed.)

19. It is desired to string upon a cord a great many uniform heavy rods of varying lengths so that when the cord is hung up with the rods dangling from it the rods will be equally spaced along the horizontal and have their lower ends on the same level. Required the shape the cord will take. (It should be noted that the shape must be known before the rods can be cut in the proper lengths to hang as desired.) The weight of the cord may be neglected.

20. A masonry arch carries a horizontal roadbed. On the assumption that the material between the arch and the roadbed is of uniform density and that each element of the arch supports the weight of the material above it, find the shape of the arch.

21. In equations (4′) the integration may be carried through in terms of quadratures if ρY is a function of y alone; and similarly in Ex. 15 the integration may be carried through if $\Phi = 0$ and ρR is a function of r alone so that the field is central. Show that the results of thus carrying through the integration are the formulas

$$x + C' = \int \frac{C\,dy}{\sqrt{\left(\int \rho Y dy\right)^2 - C^2}}, \qquad \phi + C' = \int \frac{C\,dr/r}{\sqrt{\left(\int \rho R dr\right)^2 - C^2}}.$$

22. A particle falls from rest through the air, which is assumed to offer a resistance proportional to the velocity. Solve the problem with the initial conditions $v = 0$, $x = 0$, $t = 0$. Show that as the particle falls, the velocity does not increase indefinitely, but approaches a definite limit $V = mg/k$.

23. Solve Ex. 22 with the initial conditions $v = v_0$, $x = 0$, $t = 0$, where v_0 is greater than the limiting velocity V. Show that the particle slows down as it falls.

24. A particle rises through the air, which is assumed to resist proportionally to the square of the velocity. Solve the motion.

25. Solve the problem analogous to Ex. 24 for a falling particle. Show that there is a limiting velocity $V = \sqrt{mg/k}$. If the particle were projected down with an initial velocity greater than V, it would slow down as in Ex. 23.

26. A particle falls towards a point which attracts it inversely as the square of the distance and directly as its mass. Find the relation between x and t and determine the total time T taken to reach the center. Initial conditions $v = 0$, $x = a$, $t = 0$.

$$\sqrt{\frac{2k}{a}}\,t = \frac{a}{2}\cos^{-1}\frac{2x - a}{a} + \sqrt{ax - x^2}, \qquad T = \pi k^{-\frac{1}{2}}\left(\frac{a}{2}\right)^{\frac{3}{2}}.$$

* Exercises 17–20 should be worked *ab initio* by the method by which (4) were derived, not by applying (4) directly.

27. A particle starts from the origin with a velocity V and moves in a medium which resists proportionally to the velocity. Find the relations between velocity and distance, velocity and time, and distance and time; also the limiting distance traversed.

$$v = V - kx/m, \qquad v = Ve^{-\frac{k}{m}t}, \qquad kx = mV(1 - e^{-\frac{k}{m}t}), \qquad mV/k.$$

28. Solve Ex. 27 under the assumption that the resistance varies as $\sqrt{v}$.

29. A particle falls toward a point which attracts inversely as the cube of the distance and directly as the mass. The initial conditions are $x = a$, $v = 0$, $t = 0$. Show that $x^2 = a^2 - kt^2/a^2$ and the total time of descent is $T = a^2/\sqrt{k}$.

30. A cylindrical spar buoy stands vertically in the water. The buoy is pressed down a little and released. Show that, if the resistance of the water and air be neglected, the motion is simple harmonic. Integrate and determine the constants from the initial conditions $x = 0$, $v = V$, $t = 0$, where x measures the displacement from the position of equilibrium.

31. A particle slides down a rough inclined plane. Determine the motion. Note that of the force of gravity only the component $mg \sin i$ acts down the plane, whereas the component $mg \cos i$ acts perpendicularly to the plane and develops the force $\mu mg \cos i$ of friction. Here i is the inclination of the plane and μ is the coefficient of friction.

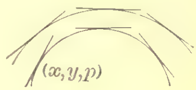
32. A bead is free to move upon a frictionless wire in the form of an inverted cycloid (vertex down). Show that the component of the weight along the tangent to the cycloid is proportional to the distance of the particle from the vertex. Hence determine the motion as simple harmonic and fix the constants of integration by the initial conditions that the particle starts from rest at the top of the cycloid.

33. Two equal weights are hanging at the end of an elastic string. One drops off. Determine completely the motion of the particle remaining.

34. One end of an elastic spring (such as is used in a spring balance) is attached rigidly to a point on a horizontal table. To the other end a particle is attached. If the particle be held at such a point that the spring is elongated by the amount a and then released, determine the motion on the assumption that the coefficient of friction between the particle and the table is μ; and discuss the possibility of different cases according as the force of friction is small or large relative to the force exerted by the spring.

85. Lineal element and differential equation. The idea of a curve as made up of the points upon it is familiar. Points, however, have no extension and therefore must be regarded not as pieces of a curve but merely as positions on it. Strictly speaking, the pieces of a curve are the elements Δs of arc; but for many purposes it is convenient to replace the complicated element Δs by a piece of the tangent to the curve at some point of the arc Δs, and from this point of view a curve is made up of an infinite number of infinitesimal elements tangent to it. This is analogous to the point of view by which a curve is regarded as made

up of an infinite number of infinitesimal chords and is intimately related to the conception of the curve as the envelope of its tangents (§ 65). A point on a curve taken with an infinitesimal portion of the tangent to the curve at that point is called a *lineal element* of the curve. These concepts and definitions are clearly equally available in two or three dimensions. For the present the curves under discussion will be plane curves and the lineal elements will therefore all lie in a plane.



To specify any particular lineal element *three coördinates* x, y, p will be used, of which the two (x, y) determine the point through which the element passes and of which the third p is the slope of the element. If a curve $f(x, y) = 0$ is given, the slope at any point may be found by differentiation,

$$p = \frac{dy}{dx} = -\frac{\partial f}{\partial x} \bigg/ \frac{\partial f}{\partial y}, \tag{6}$$

and hence the third coördinate p of the lineal elements of this particular curve is expressed in terms of the other two. If in place of one curve $f(x, y) = 0$ the whole family of curves $f(x, y) = C$, where C is an arbitrary constant, had been given, the slope p would still be found from (6), and it therefore appears that the third coördinate of the lineal elements of such a family of curves is expressible in terms of x and y.

In the more general case where the family of curves is given in the unsolved form $F(x, y, C) = 0$, the slope p is found by the same formula but it now depends apparently on C in addition to on x and y. If, however, the constant C be eliminated from the two equations

$$F(x, y, C) = 0 \quad \text{and} \quad \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} p = 0, \tag{7}$$

there will arise an equation $\Phi(x, y, p) = 0$ which connects the slope p of any curve of the family with the coördinates (x, y) of any point through which a curve of the family passes and at which the slope of that curve is p. Hence it appears that the three coördinates (x, y, p) of the lineal elements of all the curves of a family are connected by an equation $\Phi(x, y, p) = 0$, just as the coördinates (x, y, z) of the points of a surface are connected by an equation $\Phi(x, y, z) = 0$. As the equation $\Phi(x, y, z) = 0$ is called the equation of the surface, so the equation $\Phi(x, y, p) = 0$ is called the equation of the family of curves; it is, however, not the finite equation $F(x, y, C) = 0$ but the differential equation of the family, because it involves the derivative $p = dy/dx$ of y by x instead of the parameter C.

As an example of the elimination of a constant, consider the case of the parabolas

$$y^2 = Cx \quad \text{or} \quad y^2/x = C.$$

The differentiation of the equation in the second form gives at once

$$-y^2/x^2 + 2\,yp/x = 0 \quad \text{or} \quad y = 2\,xp$$

as the differential equation of the family. In the unsolved form the work is

$$2\,yp = C, \qquad y^2 = 2\,ypx, \qquad y = 2\,xp.$$

The result is, of course, the same in either case. For the family here treated it makes little difference which method is followed. As a general rule it is perhaps best to solve for the constant if the solution is simple and leads to a simple form of the function $f(x, y)$; whereas if the solution is not simple or the form of the function is complicated, it is best to differentiate first because the differentiated equation may be simpler to solve for the constant than the original equation, or because the elimination of the constant between the two equations can be conducted advantageously.

If an equation $\Phi(x, y, p) = 0$ connecting the three coördinates of the lineal element be given, the elements which satisfy the equation may be plotted much as a surface is plotted: that is, a pair of values (x, y) may be assumed and substituted in the equation, the equation may then be solved for one or more values of p, and lineal elements with these values of p may be drawn through the point (x, y). In this manner the elements through as many points as desired may be found. The detached elements are of interest and significance chiefly from the fact that they can be *assembled into curves*, — in fact, into the curves of a family $F(x, y, C) = 0$ of which the equation $\Phi(x, y, p) = 0$ is the differential equation. This is the converse of the problem treated above and requires the integration of the differential equation $\Phi(x, y, p) = 0$ for its solution. In some simple cases the assembling may be accomplished intuitively from the geometric properties implied in the equation, in other cases it follows from the integration of the equation by analytic means, in other cases it can be done only approximately and by methods of computation.

As an example of intuitively assembling the lineal elements into curves, take

$$\Phi(x, y, p) = y^2p^2 + y^2 - r^2 = 0 \quad \text{or} \quad p = \pm \frac{\sqrt{r^2 - y^2}}{y}.$$

The quantity $\sqrt{r^2 - y^2}$ may be interpreted as one leg of a right triangle of which y is the other leg and r the hypotenuse. The slope of the hypotenuse is then $\pm\, y/\sqrt{r^2 - y^2}$ according to the position of the figure, and the differential equation $\Phi(x, y, p) = 0$ states that the coördinate p of the lineal element which satisfies it is the negative reciprocal of this slope. Hence the lineal element is perpendicular to the hypotenuse. It therefore appears that the lineal elements are tangent to circles of radius r described about points of the x-axis. The equation of these circles is

$(x - C)^2 + y^2 = r^2$, and this is therefore the integral of the differential equation. The correctness of this integral may be checked by direct integration. For

$$p = \frac{dy}{dx} = \pm \frac{\sqrt{r^2 - y^2}}{y} \quad \text{or} \quad \frac{ydy}{\sqrt{r^2 - y^2}} = dx \quad \text{or} \quad \sqrt{r^2 - y^2} = x - C.$$

86. In geometric problems which relate the slope of the tangent of a curve to other lines in the figure, it is clear that not the tangent but the lineal element is the vital thing. Among such problems that of the *orthogonal trajectories* (or trajectories under any angle) of a given family of curves is of especial importance. If two families of curves are so related that the angle at which any curve of one of the families cuts any curve of the other family is a right angle, then the curves of either family are said to be the orthogonal trajectories of the curves of the other family. Hence at any point (x, y) at which two curves belonging to the different families intersect, there are two lineal elements, one belonging to each curve, which are perpendicular. As the slopes of two perpendicular lines are the negative reciprocals of each other, it follows that if the coördinates of one lineal element are (x, y, p) the coördinates of the other are $(x, y, -1/p)$; and if the coördinates of the lineal element (x, y, p) satisfy the equation $\Phi(x, y, p) = 0$, the coördinates of the orthogonal lineal element must satisfy $\Phi(x, y, -1/p) = 0$. Therefore the *rule for finding the orthogonal trajectories of the curves $F(x, y, C) = 0$ is to find first the differential equation $\Phi(x, y, p) = 0$ of the family, then to replace p by $-1/p$ to find the differential equation of the orthogonal family, and finally to integrate this equation to find the family.* It may be noted that if $F(z) = X(x, y) + iY(x, y)$ is a function of $z = x + iy$ (§ 73), the families $X(x, y) = C$ and $Y(x, y) = K$ are orthogonal.

As a problem in orthogonal trajectories find the trajectories of the semicubical parabolas $(x - C)^3 = y^2$. The differential equation of this family is found as

$$3(x - C)^2 = 2yp, \qquad x - C = (\tfrac{2}{3}yp)^{\frac{1}{2}}, \qquad (\tfrac{2}{3}yp)^{\frac{3}{2}} = y^2 \quad \text{or} \quad \tfrac{2}{3}p = y^{\frac{1}{3}}.$$

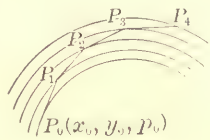
This is the differential equation of the given family. Replace p by $-1/p$ and integrate:

$$-\frac{2}{3p} = y^{\frac{1}{3}} \quad \text{or} \quad 1 + \frac{3}{2}py^{\frac{1}{3}} = 0 \quad \text{or} \quad dx + \frac{3}{2}y^{\frac{1}{3}}dy = 0, \quad \text{and} \quad x + \frac{9}{8}y^{\frac{4}{3}} = C.$$

Thus the differential equation and finite equation of the orthogonal family are found. The curves look something like parabolas with axis horizontal and vertex toward the right.

Given a differential equation $\Phi(x, y, p) = 0$ or, in solved form, $p = \phi(x, y)$; the *lineal element affords a means for obtaining graphically and numerically an approximation to the solution which passes through*

a [illegible] *point* $P_0(x_0, y_0)$. For the value p_0 of p at this point may be computed from the equation and a lineal element P_0P_1 may be drawn, the length being taken small. As the lineal element is tangent to the curve, its end point will not lie upon the curve but will depart from it by an infinitesimal of higher order. Next the slope p_1 of the lineal element which satisfies the equation and passes through P_1 may be found and the element P_1P_2 may be drawn. This element will not be tangent to the desired solution but to a solution lying near that one. Next the element P_2P_3 may be drawn, and so on. The broken line $P_0P_1P_2P_3 \cdots$ is clearly an approximation to the solution and will be a better approximation the shorter the elements P_iP_{i+1} are taken. If the radius of curvature of the solution at P_0 is not great, the curve will be bending rapidly and the elements must be taken fairly short in order to get a fair approximation; but if the radius of curvature is great, the elements need not be taken so small. (This method of approximate graphical solution indicates a method which is of value in proving by the method of limits that the equation $p = \phi(x, y)$ actually has a solution; but that matter will not be treated here.)



Let it be required to plot approximately that solution of $yp + x = 0$ which passes through (0, 1) and thus to find the ordinate for $x = 0.5$, and the area under the curve and the length of the curve to this point. Instead of assuming the lengths of the successive lineal elements, let the lengths of successive increments δx of x be taken as $\delta x = 0.1$. At the start $x_0 = 0$, $y_0 = 1$, and from $p = -x/y$ it follows that $p_0 = 0$. The increment δy of y acquired in moving along the tangent is $\delta y = p\delta x = 0$. Hence the new point of departure (x_1, y_1) is (0.1, 1) and the new slope is $p_1 = -x_1/y_1 = -0.1$. The results of the work, as it is continued, may be grouped in the table. Hence it appears that the final ordinate is $y = 0.90$. By adding up the trapezoids the area is computed as 0.48, and by finding the elements $\delta s = \sqrt{\delta x^2 + \delta y^2}$ the length is found as 0.51. Now the particular equation here treated can be integrated.

i	δx	δy	x_i	y_i	p_i
0	...	...	0.	1.00	0.
1	0.1	0.	0.1	1.00	-0.1
2	0.1	-0.01	0.2	0.99	-0.2
3	0.1	-0.02	0.3	0.97	-0.31
4	0.1	-0.03	0.4	0.94	-0.43
5	0.1	-0.04	0.5	0.90	...

$$yp + x = 0, \qquad ydy + xdx = 0, \qquad x^2 + y^2 = C, \quad \text{and hence} \quad x^2 + y^2 = 1$$

is the solution which passes through (0, 1). The ordinate, area, and length found from the curve are therefore 0.87, 0.48, 0.52 respectively. The errors in the approximate results to two places are therefore respectively 3, 0, 2 per cent. If δx had been chosen as 0.01 and four places had been kept in the computations, the errors would have been smaller.

EXERCISES

1. In the following cases eliminate the constant C to find the differential equation of the family given:

(α) $x^2 = 2Cy + C^2$, (β) $y = Cx + \sqrt{1 - C^2}$,

(γ) $x^2 - y^2 = Cx$, (δ) $y = x\tan(x + C)$,

(ϵ) $\dfrac{x^2}{a^2 - C} + \dfrac{y^2}{b^2 - C} = 1$, *Ans.* $\left(\dfrac{dy}{dx}\right)^2 + \dfrac{(x^2 - y^2) - (a^2 - b^2)}{xy}\dfrac{dy}{dx} - 1 = 0$.

2. Plot the lineal elements and intuitively assemble them into the solution:

$$(\alpha)\ yp + x = 0, \qquad (\beta)\ xp - y = 0, \qquad (\gamma)\ r\frac{d\phi}{dr} = 1.$$

Check the results by direct integration of the differential equations.

3. Lines drawn from the points $(\pm c, 0)$ to the lineal element are equally inclined to it. Show that the differential equation is that of Ex. 1 (ϵ). What are the curves?

4. The trapezoidal area under the lineal element equals the sectorial area formed by joining the origin to the extremities of the element (disregarding infinitesimals of higher order). (α) Find the differential equation and integrate. (β) Solve the same problem where the areas are equal in magnitude but opposite in sign. What are the curves?

5. Find the orthogonal trajectories of the following families. Sketch the curves.

(α) parabolas $y^2 = 2Cx$, *Ans.* ellipses $2x^2 + y^2 = C$.

(β) exponentials $y = Ce^{kx}$, *Ans.* parabolas $\frac{1}{2}ky^2 + x = C$.

(γ) circles $(x - C)^2 + y^2 = a^2$, *Ans.* tractrices.

(δ) $x^2 - y^2 = C^2$. (ϵ) $Cy^2 = x^3$. (ζ) $x^{\frac{2}{3}} + y^{\frac{2}{3}} = C^{\frac{2}{3}}$.

6. Show from the answer to Ex. 1 (ϵ) that the family is self-orthogonal and illustrate with a sketch. From the fact that the lineal element of a parabola makes equal angles with the axis and with the line drawn to the focus, derive the differential equation of all coaxial confocal parabolas and show that the family is self-orthogonal.

7. If $\Phi(x, y, p) = 0$ is the differential equation of a family, show

$$\Phi\left(x, y, \frac{p - m}{1 + mp}\right) = 0 \quad \text{and} \quad \Phi\left(x, y, \frac{p + m}{1 - mp}\right) = 0$$

are the differential equations of the family whose curves cut those of the given family at $\tan^{-1} m$. What is the difference between these two cases?

8. Show that the differential equations

$$\Phi\left(\frac{dr}{d\phi}, r, \phi\right) = 0 \quad \text{and} \quad \Phi\left(-r^2\frac{d\phi}{dr}, r, \phi\right) = 0$$

define orthogonal families in polar coördinates, and write the equation of the family which cuts the first of these at the constant angle $\tan^{-1} m$.

9. Find the orthogonal trajectories of the following families. Sketch.

(α) $r = e^{C\phi}$, (β) $r = C(1 - \cos\phi)$, (γ) $r = C\phi$, (δ) $r^2 = C^2\cos 2\phi$.

10. Recompute the approximate solution of $yp + x = 0$ under the conditions of the text but with $\delta x = 0.05$, and carry the work to three decimals.

11. Plot the approximate solution of $p = xy$ between (1, 1) and the y-axis. Take $\delta x = -0.2$. Find the ordinate, area, and length. Check by integration and comparison.

12. Plot the approximate solution of $p = -x$ through (1, 1), taking $\delta x = 0.1$ and following the curve to its intersection with the x-axis. Find also the area and the length.

13. Plot the solution of $p = \sqrt{x^2 + y^2}$ from the point (0, 1) to its intersection with the x-axis. Take $\delta x = -0.2$ and find the area and length.

14. Plot the solution of $p = s$ which starts from the origin into the first quadrant (s is the length of the arc). Take $\delta x = 0.1$ and carry the work for five steps to find the final ordinate, the area, and the length. Compare with the true integral.

87. The higher derivatives; analytic approximations. Although a differential equation $\Phi(x, y, y') = 0$ does not determine the relation between x and y without the application of some process equivalent to integration, it does afford a means of computing the higher derivatives simply by differentiation. Thus

$$\frac{d\Phi}{dx} = \frac{\partial\Phi}{\partial x} + \frac{\partial\Phi}{\partial y}y' + \frac{\partial\Phi}{\partial y'}y'' = 0$$

is an equation which may be solved for y'' as a function of x, y, y'; and y'' may therefore be expressed in terms of x and y by means of $\Phi(x, y, y') = 0$. A further differentiation gives the equation

$$\frac{d^2\Phi}{dx^2} = \frac{\partial^2\Phi}{\partial x^2} + 2\frac{\partial^2\Phi}{\partial x\partial y}y' + 2\frac{\partial^2\Phi}{\partial x\partial y'}y'' + \frac{\partial^2\Phi}{\partial y^2}y'^2 + 2\frac{\partial^2\Phi}{\partial y\partial y'}y'y''$$
$$+ \frac{\partial^2\Phi}{\partial y'^2}y''^2 + \frac{\partial\Phi}{\partial y}y'' + \frac{\partial\Phi}{\partial y'}y''' = 0,$$

which may be solved for y''' in terms of x, y, y', y''; and hence, by the preceding results, y''' is expressible as a function of x and y; and so on to all the higher derivatives. In this way any property of the integrals of $\Phi(x, y, y') = 0$ which, like the radius of curvature, is expressible in terms of the derivatives, may be found as a function of x and y.

As the differential equation $\Phi(x, y, y') = 0$ defines y' and all the higher derivatives as functions of x, y, it is clear that the values of the derivatives may be found as y_0', y_0'', y_0''', $\cdots$ at any given point (x_0, y_0). Hence it is possible to write the series

$$y = y_0 + y_0'(x - x_0) + \tfrac{1}{2}y_0''(x - x_0)^2 + \tfrac{1}{6}y_0'''(x - x_0)^3 + \cdots. \quad (8)$$

If this power series in $x - x_0$ converges, it defines y as a function of x for values of x near x_0; it is indeed *the Taylor development of the*

function y (§ 167). The convergence is assumed. Then

$$y' = y'_0 + y''_0(x - x_0) + \tfrac{1}{2} y'''_0 (x - x_0)^2 + \cdots.$$

It may be shown that the function y defined by the series actually satisfies the differential equation $\Phi(x, y, y') = 0$, that is, that

$$\Omega(x) = \Phi[x, y_0 + y'_0(x - x_0) + \tfrac{1}{2} y''_0(x - x_0)^2 + \cdots, y'_0 + y''_0(x - x_0) + \cdots] = 0$$

for all values of x near x_0. To prove this accurately, however, is beyond the scope of the present discussion; the fact may be taken for granted. Hence an analytic expansion for the integral of a differential equation has been found.

As an example of computation with higher derivatives let it be required to determine the radius of curvature of that solution of $y' = \tan(y/x)$ which passes through (1, 1). Here the slope $y'_{(1,1)}$ at (1, 1) is $\tan 1 = 1.557$. The second derivative is

$$y'' = \frac{dy'}{dx} = \frac{d}{dx}\tan\frac{y}{x} = \sec^2\frac{y}{x}\,\frac{xy' - y}{x^2}.$$

From these data the radius of curvature is found to be

$$R = \frac{(1 + y'^2)^{\frac{3}{2}}}{y''} = \sec\frac{y}{x}\,\frac{x^2}{xy' - y}, \qquad R_{(1,1)} = \sec 1\,\frac{1}{\tan 1 - 1} = 3.250.$$

The equation of the circle of curvature may also be found. For as $y''_{(1,1)}$ is positive, the curve is concave up. Hence $(1 - 3.250 \sin 1,\ 1 + 3.250 \cos 1)$ is the center of curvature; and the circle is

$$(x + 1.735)^2 + (y - 2.757)^2 = (3.250)^2.$$

As a second example let four terms of the expansion of that integral of $x \tan y' = y$ which passes through (2, 1) be found. The differential equation may be solved; then

$$\frac{dy}{dx} = \tan^{-1}\left(\frac{y}{x}\right), \qquad \frac{d^2y}{dx^2} = \frac{xy' - y}{x^2 + y^2},$$

$$\frac{d^3y}{dx^3} = \frac{(x^2 + y^2)(x - 1)y'' + (3y^2 - x^2)y' - 2xyy'^2 + 2xy}{(x^2 + y^2)^2}.$$

Now it must be noted that the problem is not wholly determinate; for y' is multiple valued and any one of the values for $\tan^{-1}\frac{1}{2}$ may be taken as the slope of a solution through (2, 1). Suppose that the angle be taken in the first quadrant; then $\tan^{-1}\frac{1}{2} = 0.462$. Substituting this in y'', we find $y''_{(2,1)} = -0.0152$; and hence may be found $y'''_{(2,1)} = 0.110$. The series for y to four terms is therefore

$$y = 1 + 0.462(x - 2) - 0.0076(x - 2)^2 + 0.018(x - 2)^3.$$

It may be noted that it is generally simpler not to express the higher derivatives in terms of x and y, but to compute each one successively from the preceding ones.

88. Picard has given a method for the integration of the equation $y' = \phi(x, y)$ by *successive approximations* which, although of the highest theoretic value and importance, is not particularly suitable to analytic

uses in finding an approximate solution. The method is this. Let the equation $y' = \phi(x, y)$ be given in solved form, and suppose (x_0, y_0) is the point through which the solution is to pass. To find the first approximation let y be held constant and equal to y_0, and integrate the equation $y' = \phi(x, y_0)$. Thus

$$dy = \phi(x, y_0)\,dx; \qquad y = y_0 + \int_{x_0}^{x} \phi(x, y_0)\,dx = f_1(x). \tag{9}$$

where it will be noticed that the constant of integration has been chosen so that the curve passes through (x_0, y_0). For the second approximation let y have the value just found, substitute this in $\phi(x, y)$, and integrate again. Then

$$y = y_0 + \int_{x_0}^{x} \phi\left[x, y_0 + \int_{x_0}^{x} \phi(x, y_0)\,dx\right]dx = f_2(x). \tag{9'}$$

With this new value for y continue as before. The successive determinations of y as a function of x actually converge toward a limiting function which is a solution of the equation and which passes through (x_0, y_0). It may be noted that at each step of the work an integration is required. The difficulty of actually performing this integration in formal practice limits the usefulness of the method in such cases. It is clear, however, that with an integrating machine such as the integraph the method could be applied as rapidly as the curves $\phi(x, f_i(x))$ could be plotted.

To see how the method works, consider the integration of $y' = x + y$ to find the integral through (1, 1). For the first approximation $y = 1$. Then

$$dy = (x + 1)\,dx, \qquad y = \tfrac{1}{2}x^2 + x + C, \qquad y = \tfrac{1}{2}x^2 + x - \tfrac{1}{2} = f_1(x).$$

From this value of y the next approximation may be found, and then still another:

$$dy = [x + (\tfrac{1}{2}x^2 + x - \tfrac{1}{2})]\,dx, \qquad y = \tfrac{1}{6}x^3 + x^2 - \tfrac{1}{2}x + \tfrac{1}{3} = f_2(x),$$
$$dy = [x + f_2(x)]\,dx, \qquad y = \tfrac{1}{24}x^4 + \tfrac{1}{3}x^3 + \tfrac{1}{4}x^2 + \tfrac{1}{3}x + \tfrac{1}{24}.$$

In this case there are no difficulties which would prevent any number of applications of the method. In fact it is evident that if y' is a polynomial in x and y, the result of any number of applications of the method will be a polynomial in x.

The method of *undetermined coefficients* may often be employed to advantage to develop the solution of a differential equation into a series. The result is of course identical with that obtained by the application of successive differentiation and Taylor's series as above; the work is sometimes shorter. Let the equation be in the form $y' = \phi(x, y)$ and assume an integral in the form

$$y = y_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + \cdots \tag{10}$$

Then $\phi(x, y)$ may also be expanded into a series, say,

$$\phi(x, y) = A_0 + A_1(x - x_0) + A_2(x - x_0)^2 + A_3(x - x_0)^3 + \cdots.$$

But by differentiating the assumed form for y we have

$$y' = a_1 + 2a_2(x - x_0) + 3a_3(x - x_0)^2 + 4a_4(x - x_0)^3 + \cdots.$$

Thus there arise two different expressions as series in $x - x_0$ for the function y', and therefore the corresponding coefficients must be equal. The resulting set of equations

$$a_1 = A_0, \quad 2a_2 = A_1, \quad 3a_3 = A_2, \quad 4a_4 = A_3, \quad \cdots \tag{11}$$

may be solved successively for the undetermined coefficients a_1, a_2, a_3, a_4, $\cdots$ which enter into the assumed expansion. This method is particularly useful when the form of the differential equation is such that some of the terms may be omitted from the assumed expansion (see Ex. 14).

As an example in the use of undetermined coefficients consider that solution of the equation $y' = \sqrt{x^2 + 3y^2}$ which passes through $(1, 1)$. The expansion will proceed according to powers of $x - 1$, and for convenience the variable may be changed to $t = x - 1$ so that

$$\frac{dy}{dt} = \sqrt{(t+1)^2 + 3y^2}, \qquad y = 1 + a_1t + a_2t^2 + a_3t^3 + a_4t^4 + \cdots$$

are the equation and the assumed expansion. One expression for y' is

$$y' = a_1 + 2a_2t + 3a_3t^2 + 4a_4t^3 + \cdots.$$

To find the other it is necessary to expand into a series in t the expression

$$y' = \sqrt{(1+t)^2 + 3(1 + a_1t + a_2t^2 + a_3t^3)^2}.$$

If this had to be done by Maclaurin's series, nothing would be gained over the method of § 87; but in this and many other cases algebraic methods and known expansions may be applied (§ 32). First square y and retain only terms up to the third power. Hence

$$y' = 2\sqrt{1 + \tfrac{1}{2}(1 + 3a_1)t + \tfrac{1}{4}(1 + 6a_2 + 3a_1^2)t^2 + \tfrac{3}{2}(a_1a_2 + a_3)t^3}.$$

Now let the quantity under the radical be called $1 + h$ and expand so that

$$y' = 2\sqrt{1+h} = 2(1 + \tfrac{1}{2}h - \tfrac{1}{8}h^2 + \tfrac{1}{16}h^3).$$

Finally raise h to the indicated powers and collect in powers of t. Then

$$y' = 2 + \tfrac{1}{2}(1 + 3a_1)\left|\begin{array}{l} t \\ \end{array}\right. \begin{array}{l|} \hfill t^2 \\ + \tfrac{1}{4}(1 + 6a_2 + 3a_1^2) \\ - \tfrac{1}{16}(1 + 3a_1)^2 \\ \end{array} \begin{array}{l|} \hfill t^3 \\ + \tfrac{3}{2}(a_1a_2 + a_3) \\ - \tfrac{1}{16}(1 + 3a_1)(1 + 6a_2 + 3a_1^2) \\ + \tfrac{1}{64}(1 + 3a_1)^3 \end{array}.$$

Hence the successive equations for determining the coefficients are $a_1 = 2$ and

$2a_2 = \frac{1}{2}(1 + 3a_1)$ or $a_2 = \frac{7}{4}$,
$3a_3 = \frac{1}{4}(1 + 6a_2 + 3a_1^2) - \frac{1}{16}(1 + 3a_1)^2$ or $a_3 = \frac{13}{16}$,
$4a_4 = \frac{3}{2}(a_1a_2 + a_3) - \frac{1}{16}(1 + 3a_1)(1 + 6a_2 + 3a_1^2) + \frac{1}{64}(1 + 3a_1)^3$ or $a_4 = \frac{111}{256}$.

Therefore to five terms the expansion desired is

$$y = 1 + 2(x - 1) + \tfrac{7}{4}(x - 1)^2 + \tfrac{13}{16}(x - 1)^3 + \tfrac{111}{256}(x - 1)^4.$$

The methods of developing a solution by Taylor's series or by undetermined coefficients apply equally well to equations of higher order. For example consider an equation of the second order in solved form $y'' = \phi(x, y, y')$ and its derivatives

$$y''' = \frac{\partial\phi}{\partial x} + \frac{\partial\phi}{\partial y}y' + \frac{\partial\phi}{\partial y'}y''$$

$$y^{\text{iv}} = \frac{\partial^2\phi}{\partial x^2} + 2\frac{\partial^2\phi}{\partial x\partial y}y' + 2\frac{\partial^2\phi}{\partial x\partial y'}y'' + \frac{\partial^2\phi}{\partial y^2}y'^2 + 2\frac{\partial^2\phi}{\partial y\partial y'}y'y'' + \frac{\partial^2\phi}{\partial y'^2}y''^2 + \frac{\partial\phi}{\partial y}y'' + \frac{\partial\phi}{\partial y'}y'''.$$

Evidently the higher derivatives of y may be obtained in terms of x, y, y'; and y itself may be written in the expanded form

$$y = y_0 + y'_0(x - x_0) + \tfrac{1}{2}y''_0(x - x_0)^2 + \tfrac{1}{6}y'''_0(x - x_0)^3 + \tfrac{1}{24}y^{\text{iv}}_0(x - x_0)^4 + \cdots, \tag{12}$$

where any desired values may be attributed to the ordinate y_0 at which the curve cuts the line $x = x_0$, and to the slope y'_0 of the curve at that point. Moreover the coefficients y''_0, y'''_0, $\cdots$ are determined in such a way that they depend on the assumed values of y_0 and y'_0. It therefore is seen that the solution (12) of the differential equation of the second order really involves two arbitrary constants, and the justification of writing it as $F(x, y, C_1, C_2) = 0$ is clear.

In following out the method of undetermined coefficients a solution of the equation would be assumed in the form

$$y = y_0 + y'_0(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + a_4(x - x_0)^4 + \cdots, \tag{13}$$

from which y' and y'' would be obtained by differentiation. Then if the series for y and y' be substituted in $y'' = \phi(x, y, y')$ and the result arranged as a series, a second expression for y'' is obtained and the comparison of the coefficients in the two series will afford a set of equations from which the successive coefficients may be found in terms of y_0 and y'_0 by solution. These results may clearly be generalized to the case of differential equations of the nth order, whereof the solutions will depend on n arbitrary constants, namely, the values assumed for y and its first $n - 1$ derivatives when $x = x_0$.

EXERCISES

1. Find the radii and circles of curvature of the solutions of the following equations at the points indicated :

(α) $y' = \sqrt{x^2 + y^2}$ at $(0, 1)$. (β) $yy' + x = 0$ at (x_0, y_0).

2. Find $y'''_{(1,\,1)} = (5\sqrt{2} - 2)/4$ if $y' = \sqrt{x^2 + y^2}$.

3. Given the equation $y^2y'^3 + xyy'^2 - yy' + x^2 = 0$ of the third degree in y' so that there will be three solutions with different slopes through any ordinary point (x, y). Find the radii of curvature of the three solutions through $(0, 1)$.

4. Find three terms in the expansion of the solution of $y' = e^{xy}$ about $(2, \frac{1}{2})$.

5. Find four terms in the expansion of the solution of $y = \log \sin xy$ about $(\frac{1}{2}\pi, 1)$.

6. Expand the solution of $y' = xy$ about $(1, y_0)$ to five terms.

7. Expand the solution of $y' = \tan (y/x)$ about $(1, 0)$ to four terms. Note that here x should be expanded in terms of y, not y in terms of x.

8. Expand two of the solutions of $y^2y'^3 + xyy'^2 - yy' + x^2 = 0$ about $(-2, 1)$ to four terms.

9. Obtain four successive approximations to the integral of $y' = xy$ through $(1, 1)$.

10. Find four successive approximations to the integral of $y' = x + y$ through $(0, y_0)$.

11. Show by successive approximations that the integral of $y' = y$ through $(0, y_0)$ is the well-known $y = y_0e^x$.

12. Carry the approximations to the solution of $y' = -x/y$ through $(0, 1)$ as far as you can integrate, and plot each approximation on the same figure with the exact integral.

13. Find by the method of undetermined coefficients the number of terms indicated in the expansions of the solutions of these differential equations about the points given :

(α) $y' = \sqrt{x + y}$, five terms, $(0, 1)$. (β) $y' = \sqrt{x + y}$, four terms, $(1, 3)$.

(γ) $y' = x + y$, n terms, $(0, y_0)$. (δ) $y' = \sqrt{x^2 + y^2}$, four terms, $(\frac{3}{5}, \frac{1}{4})$.

14. If the solution of an equation is to be expanded about $(0, y_0)$ and if the change of x into $-x$ and y' into $-y'$ does not alter the equation, the solution is necessarily symmetric with respect to the y-axis and the expansion may be assumed to contain only even powers of x. If the solution is to be expanded about $(0, 0)$ and a change of x into $-x$ and y into $-y$ does not alter the equation, the solution is symmetric with respect to the origin and the expansion may be assumed in odd powers. Obtain the expansions to four terms in the following cases and compare the labor involved in the method of undetermined coefficients with that which would be involved in performing the requisite six or seven differentiations for the application of Maclaurin's series :

(α) $y' = \dfrac{x}{\sqrt{x^2 + y^2}}$ about $(0, 2)$, (β) $y' = \sin xy$ about $(0, 1)$,

(γ) $y' = e^{xy}$ about $(0, 0)$, (δ) $y' = x^3y + xy^3$ about $(0, 0)$.

15. Expand to and including the term x^4 :

(α) $y'' = y'^2 + xy$ about $x_0 = 0$, $y_0 = a_0$, $y'_0 = a_1$ (by both methods).

(β) $xy'' + y' + y = 0$ about $x_0 = 0$, $y_0 = a_0$, $y'_0 = -a_0$ (by und. coeffs.).

CHAPTER VIII

THE COMMONER ORDINARY DIFFERENTIAL EQUATIONS

89. Integration by separating the variables. If a differential equation of the first order may be solved for y' so that

$$y' = \phi(x, y) \quad \text{or} \quad M(x, y)\,dx + N(x, y)\,dy = 0 \tag{1}$$

(where the functions ϕ, M, N are single valued or where only one specific branch of each function is selected in case the solution leads to multiple valued functions), the differential equation involves only the first power of the derivative and is said to be of the first degree. If, furthermore, it so happens that the functions ϕ, M, N are products of functions of x and functions of y so that the equation (1) takes the form

$$y' = \phi_1(x)\,\phi_2(y) \quad \text{or} \quad M_1(x)\,M_2(y)\,dx + N_1(x)\,N_2(y)\,dy = 0, \tag{2}$$

it is clear that the variables may be separated in the manner

$$\frac{dy}{\phi_2(y)} = \phi_1(x)\,dx \quad \text{or} \quad \frac{M_1(x)}{N_1(x)}\,dx + \frac{N_2(y)}{M_2(y)}\,dy = 0, \tag{2'}$$

and the integration is then immediately performed by integrating each side of the equation. It was in this way that the numerous problems considered in Chap. VII were solved.

As an example consider the equation $yy' + xy^2 = x$. Here

$$ydy + x(y^2 - 1)\,dx = 0 \quad \text{or} \quad \frac{ydy}{y^2 - 1} + xdx = 0,$$

and $$\tfrac{1}{2}\log(y^2 - 1) + \tfrac{1}{2}x^2 = C \quad \text{or} \quad (y^2 - 1)\,e^{x^2} = C.$$

The second form of the solution is found by taking the exponential of both sides of the first form after multiplying by 2.

In some differential equations (1) in which the variables are not immediately separable as above, the introduction of some change of variable, whether of the dependent or independent variable or both, may lead to a differential equation in which the new variables are separated and the integration may be accomplished. The selection of the proper change of variable is in general a matter for the exercise of ingenuity; succeeding paragraphs, however, will point out some special

types of equations for which a definite type of substitution is known to accomplish the separation.

As an example consider the equation $xdy - ydx = x\sqrt{x^2 + y^2}\,dx$, where the variables are clearly not separable without substitution. The presence of $\sqrt{x^2 + y^2}$ suggests a change to polar coördinates. The work of finding the solution is:

$$x = r\cos\theta, \quad y = r\sin\theta, \quad dx = \cos\theta dr - r\sin\theta d\theta, \quad dy = \sin\theta dr + r\cos\theta d\theta;$$

then $$xdy - ydx = r^2 d\theta, \qquad x\sqrt{x^2 + y^2}\,dx = r^2\cos\theta d(r\cos\theta).$$

Hence the differential equation may be written in the form

$$r^2 d\theta = r^2\cos\theta d(r\cos\theta) \quad \text{or} \quad \sec\theta d\theta = d(r\cos\theta),$$

and $$\log\tan\left(\tfrac{1}{2}\theta + \tfrac{1}{4}\pi\right) = r\cos\theta + C \quad \text{or} \quad \log\frac{1 + \sin\theta}{\cos\theta} = x + C.$$

Hence $$\frac{\sqrt{x^2 + y^2} + y}{x} = Ce^x \qquad \text{(on substitution for } \theta\text{)}.$$

Another change of variable which works, is to let $y = vx$. Then the work is:

$$x(vdx + xdv) - vxdx = x^2\sqrt{1 + v^2}\,dx \text{ or } dv = \sqrt{1 + v^2}\,dx.$$

Then $$\frac{dv}{\sqrt{1 + v^2}} = dx, \qquad \sinh^{-1}v = x + C, \qquad y = x\sinh(x + C).$$

This solution turns out to be shorter and the answer appears in neater form than before obtained. The great difference of form that may arise in the answer when different methods of integration are employed, is a noteworthy fact, and renders a set of answers practically worthless; two solvers may frequently waste more time in trying to get their answers reduced to a common form than each would spend in solving the problem in two ways.

90. If in the equation $y' = \phi(x, y)$ the function ϕ turns out to be $\phi(y/x)$, a function of y/x alone, that is, if the functions M and N are homogeneous functions of x, y and of the same order (§ 53), the differential equation is said to be *homogeneous* and the change of variable $y = vx$ or $x = vy$ will always result in separating the variables. The statement may be tabulated as:

$$\text{if} \qquad \frac{dy}{dx} = \phi\left(\frac{y}{x}\right), \qquad \text{substitute} \begin{cases} y = vx \\ \text{or } x = vy. \end{cases} \tag{3}$$

A sort of corollary case is given in Ex. 6 below.

As an example take $y\left(1 + e^{\frac{x}{y}}\right)dx + e^{\frac{x}{y}}(y - x)dy = 0$, of which the homogeneity is perhaps somewhat disguised. Here it is better to choose $x = vy$. Then

$$(1 + e^v)dx + e^v(1 - v)dy = 0 \quad \text{and} \quad dx = vdy + ydv.$$

Hence $$(v + e^v)dy + y(1 + e^v)dv = 0 \quad \text{or} \quad \frac{dy}{y} + \frac{1 + e^v}{v + e^v}dv = 0.$$

Hence $$\log y + \log(v + e^v) = C \quad \text{or} \quad x + ye^{\frac{x}{y}} = C.$$

If the differential equation may be arranged so that

$$\frac{dy}{dx} + X_1(x)\,y = X_2(x)\,y^n \quad \text{or} \quad \frac{dx}{dy} + Y_1(y)\,x = Y_2(y)\,x^n, \tag{4}$$

where the second form differs from the first only through the interchange of x and y and where X_1 and X_2 are functions of x alone and Y_1 and Y_2 functions of y, the equation is called a *Bernoulli equation;* and in particular if $n = 0$, so that the dependent variable does not occur on the right-hand side, the equation is called *linear*. The substitution which separates the variables in the respective cases is

$$y = ve^{-\int X_1(x)\,dx} \quad \text{or} \quad x = ve^{-\int Y_1(y)\,dy}. \tag{5}$$

To show that the separation is really accomplished and to find a general formula for the solution of any Bernoulli or linear equation, the substitution may be carried out formally. For

$$\frac{dy}{dx} = \frac{dv}{dx}e^{-\int X_1 dx} - vX_1e^{-\int X dx}.$$

The substitution of this value in the equation gives

$$\frac{dv}{dx}e^{-\int X_1 dx} = X_2 v^n e^{-n\int X_1 dx} \quad \text{or} \quad \frac{dv}{v^n} = X_2 e^{(1-n)\int X_1 dx}\,dx.$$

Hence $$v^{1-n} = (1-n)\int X_2 e^{(1-n)\int X_1 dx}\,dx, \quad \text{when} \quad n \neq 1,^*$$

or $$y^{1-n} = (1-n)\,e^{(n-1)\int X_1 dx}\left[\int X_2 e^{(1-n)\int X_1 dx}\,dx\right]. \tag{6}$$

There is an analogous form for the second form of the equation.

The equation $(x^2y^3 + xy)\,dy = dx$ may be treated by this method by writing it as

$$\frac{dx}{dy} - yx = y^3x^2 \quad \text{so that} \quad Y_1 = -y,\ Y_2 = y^3,\ n = 2.$$

Then let $$x = ve^{-\int -y\,dy} = ve^{\frac{1}{2}y^2}.$$

Then $$\frac{dx}{dy} - yx = \frac{dv}{dy}e^{\frac{1}{2}y^2} + vye^{\frac{1}{2}y^2} - yve^{\frac{1}{2}y^2} = \frac{dv}{dy}e^{\frac{1}{2}y^2}$$

and $$\frac{dv}{dy}e^{\frac{1}{2}y^2} = y^3v^2e^{y^2} \quad \text{or} \quad \frac{dv}{v^2} = y^3e^{\frac{1}{2}y^2}\,dy,$$

and $$-\frac{1}{v} = (y^2-2)\,e^{\frac{1}{2}y^2} + C \quad \text{or} \quad \frac{1}{x} = 2 - y^2 + Ce^{-\frac{1}{2}y^2}.$$

This result could have been obtained by direct substitution in the formula

$$x^{1-n} = (1-n)\,e^{(n-1)\int Y_1 dy}\left[\int Y_2 e^{(1-n)\int Y_1 dy}\,dy\right],$$

but actually to carry the method through is far more instructive.

* If $n = 1$, the variables are separated in the original equation.

EXERCISES

1. Solve the equations (variables immediately separable):

(α) $(1+x)y + (1-y)xy' = 0$, *Ans.* $xy = Ce^{y-x}$.

(β) $a(xdy + 2\,ydx) = xydy$, (γ) $\sqrt{1-x^2}\,dy + \sqrt{1-y^2}\,dx = 0$,

(δ) $(1+y^2)\,dx - \left(y + \sqrt{1+y}\right)(1+x)^{\frac{3}{2}}\,dy = 0$.

2. By various ingenious changes of variable, solve:

(α) $(x+y)^2y' = a^2$, *Ans.* $x + y = a\tan(y/a + C)$.

(β) $(x - y^2)\,dx + 2\,xydy = 0$, ($\gamma$) $xdy - ydx = (x^2+y^2)\,dx$,

(δ) $y' = x - y$, (ϵ) $yy' + y^2 + x + 1 = 0$.

3. Solve these homogeneous equations:

(α) $(2\sqrt{xy} - x)\,y' + y = 0$, *Ans.* $\sqrt{x/y} + \log y = C$.

(β) $xe^{\frac{y}{x}} + y - xy' = 0$, *Ans.* $y + x\log\log C/x = 0$.

(γ) $(x^2+y^2)\,dy = xydx$, (δ) $xy' - y = \sqrt{x^2+y^2}$.

4. Solve these Bernoulli or linear equations:

(α) $y' + y/x = y^2$, *Ans.* $xy\log Cx + 1 = 0$.

(β) $y' - y\csc x = \cos x - 1$, *Ans.* $y = \sin x + C\tan\frac{1}{2}x$.

(γ) $xy' + y = y^2\log x$, *Ans.* $y^{-1} = \log x + 1 + Cx$.

(δ) $(1+y^2)\,dx = (\tan^{-1}y - x)\,dy$, ($\epsilon$) $ydx + (ax^2y^n - 2\,x)\,dy = 0$,

(ζ) $xy' - ay = x + 1$, (η) $yy' + \frac{1}{2}y^2 = \cos x$.

5. Show that the substitution $y = vx$ always separates the variables in the homogeneous equation $y' = \phi(y/x)$ and derive the general formula for the integral.

6. Let a differential equation be reducible to the form

$$\frac{dy}{dx} = \phi\left(\frac{a_1x + b_1y + c_1}{a_2x + b_2y + c_2}\right), \qquad \begin{array}{r} a_1b_2 - a_2b_1 \neq 0, \\ \text{or} \quad a_1b_2 - a_2b_1 = 0. \end{array}$$

In case $a_1b_2 - a_2b_1 \neq 0$, the two lines $a_1x + b_1y + c_1 = 0$ and $a_2x + b_2y + c_2 = 0$ will meet in a point. Show that a transformation to this point as origin makes the new equation homogeneous and hence soluble. In case $a_1b_2 - a_2b_1 = 0$, the two lines are parallel and the substitution $z = a_2x + b_2y$ or $z = a_1x + b_1y$ will separate the variables.

7. By the method of Ex. 6 solve the equations:

(α) $(3\,y - 7\,x + 7)\,dx + (7\,y - 3\,x + 3)\,dy = 0$, *Ans.* $(y - x + 1)^2(y + x - 1)^5 = C$.

(β) $(2\,x + 3\,y - 5)\,y' + (3\,x + 2\,y - 5) = 0$, ($\gamma$) $(4\,x + 3\,y + 1)\,dx + (x + y + 1)\,dy = 0$,

(δ) $(2\,x + y) = y'(4\,x + 2\,y - 1)$, ($\epsilon$) $\dfrac{dy}{dx} = \left(\dfrac{x - y - 1}{2\,x - 2\,y + 1}\right)^2$.

8. Show that if the equation may be written as $yf(xy)\,dx + xg(xy)\,dy = 0$, where f and g are functions of the product xy, the substitution $v = xy$ will separate the variables.

9. By virtue of Ex. 8 integrate the equations:

(α) $(y + 2\,xy^2 - x^2y^3)\,dx + 2\,x^2ydy = 0$, *Ans.* $x + x^2y = C(1 - xy)$.

(β) $(y + xy^2)\,dx + (x - x^2y)\,dy = 0$, ($\gamma$) $(1 + xy)\,xy^2dx + (xy - 1)\,xdy = 0$.

10. By any method that is applicable solve the following. If more than one method is applicable, state what methods, and any apparent reasons for choosing one:

(α) $y' + y\cos x = y^n \sin 2x$, ($\beta$) $(2x^2y + 3y^3)\,dx = (x^3 + 2xy^2)\,dy$,
(γ) $(4x + 2y - 1)y' + 2x + y + 1 = 0$, ($\delta$) $yy' + xy^2 = x$,
(ϵ) $y'\sin^2 y + \sin x\cos y = \sin x$, ($\zeta$) $\sqrt{a^2 + x^2}(1 - y') = x + y$,
(η) $(x^3y^3 + x^2y^2 + xy + 1)\,y + (x^3y^3 - x^2y^2 - xy + 1)\,xy'$, ($\theta$) $y' = \sin(x - y)$,
(ι) $xydy - y^2dx = (x + y)^2 e^{-\frac{y}{x}}dx$, ($\kappa$) $(1 - y^2)\,dx = axy\,(x + 1)\,dy$.

91. Integrating factors. If the equation $Mdx + Ndy = 0$ by a suitable rearrangement of the terms can be put in the form of a sum of total differentials of certain functions $u, v, \cdots$, say

$$du + dv + \cdots = 0, \quad \text{then} \quad u + v + \cdots = C \tag{7}$$

is surely the solution of the equation. In this case the equation is called an *exact differential equation.* It frequently happens that although the equation cannot itself be so arranged, yet the equation obtained from it by multiplying through with a certain factor $\mu(x, y)$ may be so arranged. The factor $\mu(x, y)$ is then called an *integrating factor* of the given equation. Thus in the case of variables separable, an integrating factor is $1/M_2N_1$; for

$$\frac{1}{M_2N_1}\left[M_1M_2\,dx + N_1N_2\,dy\right] = \frac{M_1(x)}{N_1(x)}\,dx + \frac{N_2(y)}{M_2(y)}\,dy = 0; \tag{8}$$

and the integration is immediate. Again, the linear equation may be treated by an integrating factor. Let

$$dy + X_1ydx = X_2dx \quad \text{and} \quad \mu = e^{\int X_1dx}; \tag{9}$$

then
$$e^{\int X_1dx}dy + X_1e^{\int X_1dx}ydx = e^{\int X_1dx}X_2dx \tag{10}$$

or
$$d\left[ye^{\int X_1dx}\right] = e^{\int X_1dx}X_2dx, \quad \text{and} \quad ye^{\int X_1dx} = \int e^{\int X_1dx}X_2dx. \tag{11}$$

In the case of variables separable the use of an integrating factor is therefore implied in the process of separating the variables. In the case of the linear equation the use of the integrating factor is somewhat shorter than the use of the substitution for separating the variables. In general it is not possible to hit upon an integrating factor by inspection and not practicable to obtain an integrating factor by analysis, but the integration of an equation is so simple when the factor is known, and the equations which arise in practice so frequently do have simple integrating factors, that it is worth while to examine the equation to see if the factor cannot be determined by inspection and trial. To aid in the work, the differentials of the simpler functions such as

$$dxy = xdy + ydx, \qquad \tfrac{1}{2}\,d\,(x^2 + y^2) = xdx + ydy,$$

$$d\,\frac{y}{x} = \frac{xdy - ydx}{x^2}, \qquad d\tan^{-1}\frac{x}{y} = \frac{ydx - xdy}{x^2 + y^2}, \tag{12}$$

should be borne in mind.

Consider the equation $(x^4e^x - 2\,mxy^2)\,dx + 2\,mx^2ydy = 0$. Here the first term x^4e^xdx will be a differential of a function of x no matter what function of x may be assumed as a trial μ. With $\mu = 1/x^4$ the equation takes the form

$$e^xdx + 2\,m\left(\frac{ydy}{x^2} - \frac{y^2dx}{x^3}\right) = de^x + md\,\frac{y^2}{x^2} = 0.$$

The integral is therefore seen to be $e^x + my^2/x^2 = C$ without more ado. It may be noticed that this equation is of the Bernoulli type and that an integration by that method would be considerably longer and more tedious than this use of an integrating factor.

Again, consider $(x + y)\,dx - (x - y)\,dy = 0$ and let it be written as

$$xdx + ydy + ydx - xdy = 0; \quad \text{try} \quad \mu = 1/(x^2 + y^2);$$

then $$\frac{xdx + ydy}{x^2 + y^2} + \frac{ydx - xdy}{x^2 + y^2} = 0 \quad \text{or} \quad \frac{1}{2}\,d\log(x^2 + y^2) + d\tan^{-1}\frac{x}{y} = 0,$$

and the integral is $\log\sqrt{x^2 + y^2} + \tan^{-1}(x/y) = C$. Here the terms $xdx + ydy$ strongly suggested $x^2 + y^2$ and the known form of the differential of $\tan^{-1}(x/y)$ corroborated the idea. This equation comes under the homogeneous type, but the use of the integrating factor considerably shortens the work of integration.

92. The attempt has been to write $Mdx + Ndy$ or $\mu\,(Mdx + Ndy)$ as the sum of total differentials $du + dv + \cdots$, that is, as the differential dF of the function $u + v + \cdots$, so that the solution of the equation $Mdx + Ndy = 0$ could be obtained as $F = C$. When the expressions are complicated, the attempt may fail in practice even where it theoretically should succeed. It is therefore of importance to establish conditions under which a differential expression like $Pdx + Qdy$ shall be the total differential dF of some function, and to find a means of obtaining F when the conditions are satisfied. This will now be done.

Suppose $$Pdx + Qdy = dF = \frac{\partial F}{\partial x}\,dx + \frac{\partial F}{\partial y}\,dy; \tag{13}$$

then $$P = \frac{\partial F}{\partial x}, \qquad Q = \frac{\partial F}{\partial y}, \qquad \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} = \frac{\partial^2 F}{\partial x\,\partial y}.$$

Hence if $Pdx + Qdy$ is a total differential dF, it follows (as in § 52) that the relation $P'_y = Q'_x$ must hold. Now conversely if this relation does hold, it may be shown that $Pdx + Qdy$ is the total differential of a function, and that this function is

$$F = \int_{x_0}^{x} P(x, y)\,dx + \int Q(x_0, y)\,dy$$

or
$$F = \int_{y_0}^{y} Q(x, y)\,dy + \int P(x, y_0)\,dx, \tag{14}$$

where the fixed value x_0 or y_0 will naturally be so chosen as to simplify the integrations as much as possible.

To show that these expressions may be taken as F it is merely necessary to compute their derivatives for identification with P and Q. Now

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}\int_{x_0}^{x} P(x, y)\,dx + \frac{\partial}{\partial x}\int Q(x_0, y)\,dy = P(x, y),$$

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}\int_{x_0}^{x} P(x, y)\,dx + \frac{\partial}{\partial y}\int Q(x_0, y)\,dy = \frac{\partial}{\partial y}\int P dx + Q(x_0, y).$$

These differentiations, applied to the first form of F, require only the fact that the derivative of an integral is the integrand. The first turns out satisfactorily. The second must be simplified by interchanging the order of differentiation by y and integration by x (Leibniz's Rule, § 119) and by use of the fundamental hypothesis that $P'_y = Q'_x$.

$$\frac{\partial}{\partial y}\int_{x_0}^{x} P dx + Q(x_0, y) = \int_{x_0}^{x} \frac{\partial P}{\partial y}\,dx + Q(x_0, y)$$

$$= \int_{x_0}^{x} \frac{\partial Q}{\partial x}\,dx + Q(x_0, y) = Q(x, y)\Big|_{x_0}^{x} + Q(x_0, y) = Q(x, y).$$

The identity of P and Q with the derivatives of F is therefore established. The second form of F would be treated similarly.

Show that $(x^2 + \log y)\,dx + x/y dy = 0$ is an exact differential equation and obtain the solution. Here it is first necessary to apply the test $P'_y = Q'_x$. Now

$$\frac{\partial}{\partial y}(x^2 + \log y) = \frac{1}{y} \quad \text{and} \quad \frac{\partial}{\partial x}\frac{x}{y} = \frac{1}{y}.$$

Hence the test is satisfied and the integral is obtained by applying the formula:

$$\int_0^x (x^2 + \log y)\,dx + \int \frac{0}{y}\,dy = \frac{1}{3}x^3 + x\log y = C$$

or
$$\int_1^y \frac{x}{y}\,dy + \int (x^2 + \log 1)\,dx = x\log y + \frac{1}{3}x^3 = C.$$

It should be noticed that the choice of $x_0 = 0$ simplifies the integration in the first case because the substitution of the lower limit 0 is easy and because the second integral vanishes. The choice of $y_0 = 1$ introduces corresponding simplifications in the second case.

Derive the *partial differential equation which any integrating factor of the differential equation $Mdx + Ndy = 0$ must satisfy.* If μ is an integrating factor, then

$$\mu M dx + \mu N dy = dF \quad \text{and} \quad \frac{\partial \mu M}{\partial y} = \frac{\partial \mu N}{\partial x}.$$

Hence
$$M\frac{\partial \mu}{\partial y} - N\frac{\partial \mu}{\partial x} = \mu\left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) \tag{15}$$

is the desired equation. To determine the integrating factor by solving this equation would in general be as difficult as solving the original equation; in some special cases, however, this equation is useful in determining μ.

93. It is now convenient to tabulate a list of different types of differential equations for which an integrating factor of a standard form can be given. With the knowledge of the factor, the equations may then be integrated by (14) or by inspection.

EQUATION $Mdx + Ndy = 0$:	FACTOR μ:
I. Homogeneous $Mdx + Ndy = 0$,	$\dfrac{1}{Mx + Ny}$.
II. Bernoulli $dy + X_1 y dx = X_2 y^n dx$,	$y^{-n} e^{(1-n)\int X_1 dx}$.
III. $M = yf(xy)$, $N = xg(xy)$,	$\dfrac{1}{Mx - Ny}$.
IV. If $\dfrac{\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}}{N} = f(x)$,	$e^{\int f(x)dx}$.
V. If $\dfrac{\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}}{M} = f(y)$,	$e^{\int f(y)dy}$.
VI. Type $x^\alpha y^\beta(mydx + nxdy) = 0$,	$\begin{cases} x^{km-1-\alpha} y^{kn-1-\beta}, \\ k \text{ arbitrary.} \end{cases}$
VII. $x^\alpha y^\beta(mydx + nxdy) + x^\gamma y^\delta(pydx + qxdy) = 0$,	$\begin{cases} x^{km-1-\alpha} y^{kn-1-\beta}, \\ k \text{ determined.} \end{cases}$

The use of the integrating factor often is simpler than the substitution $y = vx$ in the homogeneous equation. It is practically identical with the substitution in the Bernoulli type. In the third type it is often shorter than the substitution. The remaining types have had no substitution indicated for them. The proofs that the assigned forms of the factor are right are given in the examples below or are left as exercises.

To show that $\mu = (Mx + Ny)^{-1}$ is an integrating factor for the homogeneous case, it is possible simply to substitute in the equation (15), which μ must satisfy, and show that the equation actually holds by virtue of the fact that M and N are

homogeneous of the same degree, — this fact being used to simplify the result by Euler's Formula (30) of § 53. But it is easier to proceed directly to show

$$\frac{\partial}{\partial y}\frac{M}{Mx+Ny}=\frac{\partial}{\partial x}\left(\frac{N}{Mx+Ny}\right) \quad\text{or}\quad \frac{\partial}{\partial y}\left(\frac{1}{x}\frac{1}{1+\phi}\right)=\frac{\partial}{\partial x}\left(\frac{1}{y}\frac{\phi}{1+\phi}\right),\quad\text{where}\quad \phi=\frac{Ny}{Mx}.$$

Owing to the homogeneity, ϕ is a function of y/x alone. Differentiate.

$$\frac{\partial}{\partial y}\left(\frac{1}{x}\frac{1}{1+\phi}\right)=-\frac{1}{x}\frac{\phi'}{(1+\phi)^2}\frac{1}{x}=\frac{1}{y}\frac{\phi'}{(1+\phi)^2}\cdot\frac{-y}{x^2}=\frac{\partial}{\partial x}\left(\frac{1}{y}\frac{\phi}{1+\phi}\right).$$

As this is an evident identity, the theorem is proved.

To find the condition that the integrating factor may be a function of x only and to find the factor when the condition is satisfied, the equation (15) which μ satisfies may be put in the more compact form by dividing by μ.

$$M\frac{1}{\mu}\frac{\partial\mu}{\partial y}-N\frac{1}{\mu}\frac{\partial\mu}{\partial x}=\frac{\partial N}{\partial x}-\frac{\partial M}{\partial y}\quad\text{or}\quad M\frac{\partial\log\mu}{\partial y}-N\frac{\partial\log\mu}{\partial x}=\frac{\partial N}{\partial x}-\frac{\partial M}{\partial y}. \tag{15'}$$

Now if μ (and hence $\log\mu$) is a function of x alone, the first term vanishes and

$$\frac{d\log\mu}{dx}=\frac{M'_y-N'_x}{N}=f(x)\quad\text{or}\quad \log\mu=\int f(x)\,dx.$$

This establishes the rule of type IV above and further shows that in no other case can μ be a function of x alone. The treatment of type V is clearly analogous.

Integrate the equation $x^4y(3\,ydx+2\,xdy)+x^2(4\,ydx+3\,xdy)=0$. This is of type VII; an integrating factor of the form $\mu=x^\rho y^\sigma$ will be assumed and the exponents ρ, σ will be determined so as to satisfy the condition that the equation be an exact differential. Here

$$P=\mu M=3\,x^{\rho+4}y^{\sigma+2}+4\,x^{\rho+2}y^{\sigma+1},\qquad Q=\mu N=2\,x^{\rho+5}y^{\sigma+1}+3\,x^{\rho+3}y^{\sigma}.$$

Then
$$\begin{aligned}P'_y&=3(\sigma+2)x^{\rho+4}y^{\sigma+1}+4(\sigma+1)x^{\rho+2}y^{\sigma}\\&=2(\rho+5)x^{\rho+4}y^{\sigma+1}+3(\rho+3)x^{\rho+2}y^{\sigma}=Q'_x.\end{aligned}$$

Hence if
$$3(\sigma+2)=2(\rho+5)\quad\text{and}\quad 4(\sigma+1)=3(\rho+3),$$

the relation $P'_y=Q'_x$ will hold. This gives $\sigma=2$, $\rho=1$. Hence $\mu=xy^2$,

and
$$\int_0^x(3\,x^5y^4+4\,x^3y^3)\,dx+\int 0dy=\tfrac{1}{2}x^6y^4+x^4y^3=C$$

is the solution. The work might be shortened a trifle by dividing through in the first place by x^2. Moreover the integration can be performed at sight without the use of (14).

94. Several of the most important facts relative to integrating factors and solutions of $Mdx+Ndy=0$ will now be stated as theorems and the proofs will be indicated below.

1. If an integrating factor is known, the corresponding solution may be found; and conversely if the solution is known, the corresponding integrating factor may be found. Hence the existence of either implies the existence of the other.

2. If $F=C$ and $G=C$ are two solutions of the equation, either must be a function of the other, as $G=\Phi(F)$; and any function of either is

a solution. If μ and ν are two integrating factors of the equation, the ratio μ/ν is either constant or a solution of the equation; and the product of μ by any function of a solution, as $\mu\Phi(F)$, is an integrating factor of the equation.

3. The normal derivative dF/dn of a solution obtained from the factor μ is the product $\mu\sqrt{M^2+N^2}$ (see § 48).

It has already been seen that if an integrating factor μ is known, the corresponding solution $F = C$ may be found by (14). Now if the solution is known, the equation

$$dF = F'_x dx + F'_y dy = \mu(Mdx + Ndy) \quad \text{gives} \quad F'_x = \mu M,\ F'_y = \mu N;$$

and hence μ may be found from either of these equations as the quotient of a derivative of F by a coefficient of the differential equation. The statement 1 is therefore proved. It may be remarked that the discussion of approximate solutions to differential equations (§§ 86–88), combined with the theory of limits (beyond the scope of this text), affords a demonstration that any equation $Mdx + Ndy = 0$, where M and N satisfy certain restrictive conditions, has a solution; and hence it may be inferred that such an equation has an integrating factor.

If μ be eliminated from the relations $F'_x = \mu M$, $F'_y = \mu N$ found above, it is seen that

$$MF'_y - NF'_x = 0, \quad \text{and similarly,} \quad MG'_y - NG'_x = 0, \tag{16}$$

are the conditions that F and G should be solutions of the differential equation. Now these are two simultaneous homogeneous equations of the first degree in M and N. If M and N are eliminated from them, there results the equation

$$F'_y G'_x - F'_x G'_y = 0 \quad \text{or} \quad \begin{vmatrix} F'_x & F'_y \\ G'_x & G'_y \end{vmatrix} = J(F, G) = 0, \tag{16$'$}$$

which shows (§ 62) that F and G are functionally related as required. To show that any function $\Phi(F)$ is a solution, consider the equation

$$M\Phi'_y - N\Phi'_x = (MF'_y - NF'_x)\Phi'.$$

As F is a solution, the expression $MF'_y - NF'_x$ vanishes by (16), and hence $M\Phi'_y - N\Phi'_x$ also vanishes, and Φ is a solution of the equation as is desired. The first half of 2 is proved.

Next, if μ and ν are two integrating factors, equation (15′) gives

$$M\frac{\partial \log\mu}{\partial y} - N\frac{\partial \log\mu}{\partial x} = M\frac{\partial \log\nu}{\partial y} - N\frac{\partial \log\nu}{\partial x} \quad \text{or} \quad M\frac{\partial \log\mu/\nu}{\partial y} - N\frac{\partial \log\mu/\nu}{\partial x} = 0.$$

On comparing with (16) it then appears that $\log(\mu/\nu)$ must be a solution of the equation and hence μ/ν itself must be a solution. The inference, however, would not hold if μ/ν reduced to a constant. Finally if μ is an integrating factor leading to the solution $F = C$, then

$$dF = \mu(Mdx + Ndy), \quad \text{and hence} \quad \mu\Phi(F)(Mdx + Ndy) = d\int\Phi(F)\,dF.$$

It therefore appears that the factor $\mu\Phi(F)$ makes the equation an exact differential and must be an integrating factor. Statement 2 is therefore wholly proved.

The third proposition is proved simply by differentiation and substitution. For

$$\frac{dF}{dn} = \frac{\partial F}{\partial x}\frac{dx}{dn} + \frac{\partial F}{\partial y}\frac{dy}{dn} = \mu M\frac{dx}{dn} + \mu N\frac{dy}{dn}.$$

And if τ denotes the inclination of the curve $F = C$, it follows that

$$\tan\tau = \frac{dy}{dx} = -\frac{M}{N}, \qquad \sin\tau = \frac{dy}{dn} = \frac{N}{\sqrt{M^2+N^2}}, \qquad -\cos\tau = \frac{dx}{dn} = \frac{M}{\sqrt{M^2+N^2}}.$$

Hence $dF/dn = \mu\sqrt{M^2+N^2}$ and the proposition is proved.

EXERCISES

1. Find the integrating factor by inspection and integrate:

(α) $xdy - ydx = (x^2+y^2)dx$, (β) $(y^2 - xy)dx + x^2dy = 0$,
(γ) $ydx - xdy + \log x dx = 0$, ($\delta$) $y(2xy + e^x)dx - e^xdy = 0$,
(ϵ) $(1+xy)ydx + (1-xy)xdy = 0$, ($\zeta$) $(x - y^2)dx + 2xydy = 0$,
(η) $(xy^2 + y)dx - xdy = 0$, (θ) $a(xdy + 2ydx) = xydy$,
(ι) $(x^2+y^2)(xdx + ydy) + \sqrt{1+(x^2+y^2)}\,(ydx - xdy) = 0$,
(κ) $x^2ydx - (x^3+y^3)dy = 0$, (λ) $xdy - ydx = x\sqrt{x^2-y^2}\,dy$.

2. Integrate these linear equations with an integrating factor:

(α) $y' + ay = \sin bx$, (β) $y' + y\cot x = \sec x$,
(γ) $(x+1)y' - 2y = (x+1)^4$, (δ) $(1+x^2)y' + y = e^{\tan^{-1}x}$,

and (β), (δ), (ζ) of Ex. 4, p. 206.

3. Show that the expression given under II, p. 210, is an integrating factor for the Bernoulli equation, and integrate the following equations by that method:

(α) $y' - y\tan x = y^4\sec x$, (β) $3y^2y' + y^3 - x - 1$,
(γ) $y' + y\cos x = y^n\sin 2x$, ($\delta$) $dx + 2xydy = 2ax^3y^3dy$,

and (α), (γ), (ϵ), (η) of Ex. 4, p. 206.

4. Show the following are exact differential equations and integrate:

(α) $(3x^2+6xy^2)dx + (6x^2y+4y^2)dy = 0$, ($\beta$) $\sin x\cos y dx + \cos x\sin y dy = 0$,
(γ) $(6x-2y+1)dx + (2y-2x-3)dy = 0$, ($\delta$) $(x^3+3xy^2)dx + (y^3+3x^2y)dy = 0$,
(ϵ) $\dfrac{2xy+1}{y}dx + \dfrac{y-x}{y^2}dy = 0$, ($\zeta$) $\left(1+e^{\frac{x}{y}}\right)dx + e^{\frac{x}{y}}\left(1-\dfrac{x}{y}\right)dy = 0$,
(η) $e^x(x^2+y^2+2x)dx + 2ye^xdy = 0$, ($\theta$) $(y\sin x - 1)dx + (y - \cos x)dy = 0$.

5. Show that $(Mx - Ny)^{-1}$ is an integrating factor for type III. Determine the integrating factors of the following equations, thus render them exact, and integrate:

(α) $(y+x)dx + xdy = 0$, (β) $(y^2 - xy)dx + x^2dy = 0$,
(γ) $(x^2+y^2)dx - 2xydy = 0$, (δ) $(x^2y^2 + xy)ydx + (x^2y^2-1)xdy = 0$,
(ϵ) $(\sqrt{xy}-1)xdy - (\sqrt{xy}+1)ydx = 0$, ($\zeta$) $x^3dx + (3x^2y + 2y^3)dy = 0$,

and Exs. 3 and 9, p. 206.

6. Show that the factor given for type VI is right, and that the form given for type VII is right if k satisfies $k(qm - pn) = q(\alpha-\gamma) - p(\beta-\delta)$.

7. Integrate the following equations of types IV–VII :

(α) $(y^4 + 2y)dx + (xy^3 + 2y^4 - 4x)dy = 0$, ($\beta$) $(x^2 + y^2 + 1)dx - 2xydy = 0$.
(γ) $(3x^2 + 6xy + 3y^2)dx + (2x^2 + 3xy)dy = 0$, ($\delta$) $(2x^2y^2 + y) - (x^3y - 3x)y' = 0$.
(ϵ) $(2x^2y - 3y^4)dx + (3x^3 + 2xy^3)dy = 0$,
(ζ) $(2 - y')\sin(3x - 2y) + y'\sin(x - 2y) = 0$.

8. By virtue of proposition 2 above, it follows that if an equation is exact and homogeneous, or exact and has the variables separable, or homogeneous and under types IV–VII, so that two different integrating factors may be obtained, the solution of the equation may be obtained without integration. Apply this to finding the solutions of Ex. 4 (β), (δ), (γ) ; Ex. 5 (α), (γ).

9. Discuss the apparent exceptions to the rules for types I, III, VII, that is, when $Mx + Ny = 0$ or $Mx - Ny = 0$ or $qm - pn = 0$.

10. Consider this rule for integrating $Mdx + Ndy = 0$ when the equation is known to be exact : Integrate Mdx regarding y as constant, differentiate the result regarding y as variable, and subtract from N ; then integrate the difference with respect to y. In symbols,

$$C = \int (Mdx + Ndy) = \int Mdx + \int \left(N - \frac{\partial}{\partial y}\int Mdx\right)dy.$$

Apply this instead of (14) to Ex. 4. Observe that in no case should either this formula or (14) be applied when the integral is obtainable by inspection.

95. Linear equations with constant coefficients. The type

$$a_0\frac{d^ny}{dx^n} + a_1\frac{d^{n-1}y}{dx^{n-1}} + \cdots + a_{n-1}\frac{dy}{dx} + a_ny = X(x) \tag{17}$$

of differential equation of the nth order which is of the first degree in y and its derivatives is called a *linear* equation. For the present only the case where the coefficients $a_0, a_1, \cdots, a_{n-1}, a_n$ are constant will be treated, and for convenience it will be assumed that the equation has been divided through by a_0 so that the coefficient of the highest derivative is 1. Then if differentiation be denoted by D, the equation may be written *symbolically* as

$$(D^n + a_1D^{n-1} + \cdots + a_{n-1}D + a_n)y = X, \tag{17'}$$

where the symbol D combined with constants follows many of the laws of ordinary algebraic quantities (see § 70).

The simplest equation would be of the first order. Here

$$\frac{dy}{dx} - a_1y = X \quad \text{and} \quad y = e^{a_1x}\int e^{-a_1x}Xdx, \tag{18}$$

as may be seen by reference to (11) or (6). Now if $D - a_1$ be treated as an algebraic symbol, the solution may be indicated as

$$(D - a_1)y = X \quad \text{and} \quad y = \frac{1}{D - a_1}X. \tag{18'}$$

where the operator $(D-a_1)^{-1}$ is the *inverse* of $D-a_1$. The solution which has just been obtained shows that the interpretation which must be assigned to the inverse operator is

$$\frac{1}{D-a_1}(*) = e^{a_1x}\int e^{-a_1x}(*)\,dx, \tag{19}$$

where $(*)$ denotes the function of x upon which it operates. That the integrating operator is the inverse of $D-a_1$ may be proved by direct differentiation (see Ex. 7, p. 152).

This operational method may at once be extended to obtain the solution of equations of higher order. For consider

$$\frac{d^2y}{dx^2} + a_1\frac{dy}{dx} + a_2y = X \quad \text{or} \quad (D^2 + a_1D + a_2)\,y = X. \tag{20}$$

Let α_1 and α_2 be the roots of the equation $D^2 + a_1D + a_2 = 0$ so that the differential equation may be written in the form

$$[D^2 - (\alpha_1+\alpha_2)\,D + \alpha_1\alpha_2]\,y = X \quad \text{or} \quad (D-\alpha_1)(D-\alpha_2)\,y = X. \tag{20$'$}$$

The solution may now be evaluated by a succession of steps as

$$(D-\alpha_2)\,y = \frac{1}{D-\alpha_1}X = e^{\alpha_1x}\int e^{-\alpha_1x}X\,dx,$$

$$y = \frac{1}{D-\alpha_2}\left[\frac{1}{D-\alpha_1}X\right] = e^{\alpha_2x}\int e^{-\alpha_2x}\left[e^{\alpha_1x}\int e^{-\alpha_1x}X\,dx\right]dx$$

or

$$y = e^{\alpha_2x}\int e^{(\alpha_1-\alpha_2)x}\left[\int e^{-\alpha_1x}X\,dx\right]dx. \tag{20$''$}$$

The solution of the equation is thus reduced to quadratures.

The extension of the method to an equation of any order is immediate. The first step in the solution is to solve the equation

$$D^n + a_1D^{n-1} + \cdots + a_{n-1}D + a_n = 0$$

so that the differential equation may be written in the form

$$(D-\alpha_1)(D-\alpha_2)\cdots(D-\alpha_{n-1})(D-\alpha_n)\,y = X; \tag{17$''$}$$

whereupon the solution is comprised in the formula

$$y = e^{\alpha_nx}\int e^{(\alpha_{n-1}-\alpha_n)x}\int\cdots\int e^{(\alpha_1-\alpha_2)x}\int e^{-\alpha_1x}X\,(dx)^n, \tag{17$'''$}$$

where the successive integrations are to be performed by beginning upon the extreme right and working toward the left. Moreover, it appears that if the operators $D-\alpha_n$, $D-\alpha_{n-1}, \cdots, D-\alpha_2$, $D-\alpha_1$ were successively applied to this value of y, they would undo the work here

done and lead back to the original equation. As n integrations are required, there will occur n arbitrary constants of integration in the answer for y.

As an example consider the equation $(D^3 - 4D)y = x^2$. Here the roots of the algebraic equation $D^3 - 4D = 0$ are $0, 2, -2$, and the solution for y is

$$y = \frac{1}{D}\frac{1}{D-2}\frac{1}{D+2}x^2 = \int e^{2x}\int e^{-2x}e^{-2x}\int e^{2x}x^2(dx)^3.$$

The successive integrations are very simple by means of a table. Then

$$\int e^{2x}x^2dx = \tfrac{1}{2}x^2e^{2x} - \tfrac{1}{2}xe^{2x} + \tfrac{1}{4}e^{2x} + C_1,$$

$$\begin{aligned}\int e^{-4x}\int e^{2x}x^2(dx)^2 &= \int(\tfrac{1}{2}x^2e^{-2x} - \tfrac{1}{2}xe^{-2x} + \tfrac{1}{4}e^{-2x} + C_1e^{-4x})\,dx\\ &= -\tfrac{1}{4}x^2e^{-2x} - \tfrac{1}{8}e^{-2x} + C_1e^{-4x} + C_2,\\ y = \int e^{2x}\int e^{-4x}\int e^{2x}x^2(dx)^3 &= \int(-\tfrac{1}{4}x^2 - \tfrac{1}{8} + C_1e^{-2x} + C_2e^{2x})\,dx\\ &= -\tfrac{1}{12}x^3 - \tfrac{1}{8}x + C_1e^{-2x} + C_2e^{2x} + C_3.\end{aligned}$$

This is the solution. It may be noted that in integrating a term like C_1e^{-4x} the result may be written as C_1e^{-4x}, for the reason that C_1 is arbitrary anyhow; and, moreover, if the integration had introduced any terms such as $2e^{-2x}$, $\frac{1}{2}e^{2x}$, 5, these could be combined with the terms C_1e^{-2x}, C_2e^{2x}, C_3 to simplify the form of the results.

In case the roots are imaginary the procedure is the same. Consider

$$\frac{d^2y}{dx^2} + y = \sin x \quad\text{or}\quad (D^2+1)y = \sin x \quad\text{or}\quad (D+i)(D-i)y = \sin x.$$

Then $$y = \frac{1}{D-i}\frac{1}{D+i}\sin x = e^{ix}\int e^{-2ix}\int e^{ix}\sin x\,(dx)^2, \qquad i = \sqrt{-1}.$$

The formula for $\int e^{ax}\sin bx dx$, as given in the tables, is not applicable when $a^2 + b^2 = 0$, as is the case here, because the denominator vanishes. It therefore becomes expedient to write $\sin x$ in terms of exponentials. Then

$$y = e^{ix}\int e^{-2ix}\int e^{ix}\frac{e^{ix}-e^{-ix}}{2i}(dx)^2; \quad \text{for}\quad \sin x = \frac{e^{ix}-e^{-ix}}{2i}.$$

Now $$\begin{aligned}\frac{1}{2i}e^{ix}\int e^{-2ix}\int(e^{2ix}-1)(dx)^2 &= \frac{1}{2i}e^{ix}\int e^{-2ix}\left[\frac{1}{2i}e^{2ix} - x + C_1\right]dx\\ &= \frac{1}{2i}e^{ix}\left[\frac{1}{2i}x + \frac{1}{2i}e^{-2ix}x - \frac{1}{4}e^{-2ix} + C_1e^{-2ix} + C_2\right]\\ &= -\frac{x}{2}\frac{e^{ix}+e^{-ix}}{2} + C_1e^{-ix} + C_2e^{ix}.\end{aligned}$$

Now $$C_1e^{-ix} + C_2e^{ix} = (C_2 + C_1)\frac{e^{ix}+e^{-ix}}{2} + (C_2 - C_1)\,i\,\frac{e^{ix}-e^{-ix}}{2i}.$$

Hence this expression may be written as $C_1\cos x + C_2\sin x$, and then

$$y = -\tfrac{1}{2}x\cos x + C_1\cos x + C_2\sin x.$$

The solution of such equations as these gives excellent opportunity to cultivate the art of manipulating trigonometric functions through exponentials (§ 74).

96. The general method of solution given above may be considerably simplified in case the function $X(x)$ has certain special forms. In the first place suppose $X = 0$, and let the equation be $P(D)y = 0$, where $P(D)$ denotes the symbolic polynomial of the nth degree in D. Suppose the roots of $P(D) = 0$ are $\alpha_1, \alpha_2, \cdots, \alpha_k$ and their respective multiplicities are $m_1, m_2, \cdots, m_k$, so that

$$(D - \alpha_k)^{m_k} \cdots (D - \alpha_2)^{m_2}(D - \alpha_1)^{m_1} y = 0$$

is the form of the differential equation. Now, as above, if

$$(D - \alpha_1)^{m_1} y = 0, \quad \text{then} \quad y = \frac{1}{(D - \alpha_1)^{m_1}} 0 = e^{\alpha_1 x} \int \cdots \int 0 \, (dx)^{m_1}.$$

Hence $$y = e^{\alpha_1 x}(C_1 + C_2 x + C_3 x^2 + \cdots + C_{m_1} x^{m_1 - 1})$$

is annihilated by the application of the operator $(D - \alpha_1)^{m_1}$, and therefore by the application of the whole operator $P(D)$, and must be a solution of the equation. As the factors in $P(D)$ may be written so that any one of them, as $(D - \alpha_i)^{m_i}$, comes last, it follows that to each factor $(D - \alpha_i)^{m_i}$ will correspond a solution

$$y_i = e^{\alpha_i x}(C_{i1} + C_{i2} x + \cdots + C_{im_i} x^{m_i - 1}), \qquad P(D) y_i = 0,$$

of the equation. Moreover the sum of all these solutions,

$$y = \sum_{i=1}^{i=k} e^{\alpha_i x}(C_{i1} + C_{i2} x + \cdots + C_{im_i} x^{m_i - 1}), \tag{21}$$

will be a solution of the equation; for in applying $P(D)$ to y,

$$P(D)y = P(D)y_1 + P(D)y_2 + \cdots + P(D)y_k = 0.$$

Hence the general rule may be stated that: *The solution of the differential equation $P(D)y = 0$ of the nth order may be found by multiplying each $e^{\alpha x}$ by a polynomial of $(m-1)$st degree in x (where α is a root of the equation $P(D) = 0$ of multiplicity m and where the coefficients of the polynomial are arbitrary) and adding the results.* Two observations may be made. First, the solution thus found contains n arbitrary constants and may therefore be considered as the general solution; and second, if there are imaginary roots for $P(D) = 0$, *the exponentials arising from the pure imaginary parts of the roots may be converted into trigonometric functions.*

As an example take $(D^4 - 2D^3 + D^2)y = 0$. The roots are 1, 1, 0, 0. Hence the solution is
$$y = e^x(C_1 + C_2 x) + (C_3 + C_4 x).$$

Again if $(D^4 + 4)y = 0$, the roots of $D^4 + 4 = 0$ are $\pm 1 \pm i$ and the solution is
$$y = C_1 e^{(1+i)x} + C_2 e^{(1-i)x} + C_3 e^{(-1+i)x} + C_4 e^{(-1-i)x}$$

or
$$y = e^x(C_1e^{ix} + C_2e^{-ix}) + e^{-x}(C_3e^{ix} + C_4e^{-ix})$$
$$= e^x(C_1 \cos x + C_2 \sin x) + e^{-x}(C_3 \cos x + C_4 \sin x),$$
where the new C's are not identical with the old C's. Another form is
$$y = e^x A \cos(x + \gamma) + e^{-x} B \cos(x + \delta),$$
where γ and δ, A and B, are arbitrary constants. For
$$C_1 \cos x + C_2 \sin x = \sqrt{C_1^2 + C_2^2}\left[\frac{C_1}{\sqrt{C_1^2 + C_2^2}} \cos x + \frac{C_2}{\sqrt{C_1^2 + C_2^2}} \sin x\right],$$
and if $\gamma = \tan^{-1}\left(-\frac{C_2}{C_1}\right)$, then $C_1 \cos x + C_2 \sin x = \sqrt{C_1^2 + C_2^2} \cos(x + \gamma)$.

Next if X is not zero but *if any one solution I can be found so that $P(D)\,I = X$, then a solution containing n arbitrary constants may be found by adding to I the solution of $P(D)y = 0$.* For if
$$P(D)\,I = X \quad \text{and} \quad P(D)\,y = 0, \quad \text{then} \quad P(D)(I + y) = X.$$
It therefore remains to devise means for finding one solution I. This solution I may be found by the long method of (17'''), where the integration may be shortened by omitting the constants of integration since only one, and not the general, value of the solution is needed. In the most important cases which arise in practice there are, however, some very short cuts to the solution I. The solution I of $P(D)y = X$ is called the *particular integral* of the equation and the general solution of $P(D)\,y = 0$ is called the *complementary function* for the equation $P(D)\,y = X$.

Suppose that X *is a polynomial in x.* Solve symbolically, arrange $P(D)$ in ascending powers of D, and divide out to powers of D equal to the order of the polynomial X. Then
$$P(D)\,I = X, \qquad I = \frac{1}{P(D)} X = \left[Q(D) + \frac{R(D)}{P(D)}\right] X, \tag{22}$$
where the remainder $R(D)$ is of *higher* order in D than X in x. Then
$$P(D)\,I = P(D)\,Q(D)\,X + R(D)\,X, \qquad R(D)\,X = 0.$$
Hence $Q(D)x$ may be taken as I, since $P(D)\,Q(D)\,X = P(D)\,I = X$. By this method the solution I may be found, when X is a polynomial, *as rapidly as $P(D)$ can be divided into* 1; the solution of $P(D)\,y = 0$ may be written down by (21); and the sum of I and this will be the required solution of $P(D)\,y = X$ containing n constants.

As an example consider $(D^3 + 4D^2 + 3D)\,y = x^2$. The work is as follows:
$$I = \frac{1}{3D + 4D^2 + D^3} x^2 = \frac{1}{D}\frac{1}{3 + 4D + D^2} x^2 = \frac{1}{D}\left[\frac{1}{3} - \frac{4}{9}D + \frac{13}{27}D^2 + \frac{R(D)}{P(D)}\right] x^2.$$

Hence $$I = Q(D)x^2 = \frac{1}{D}\left(\frac{1}{3} - \frac{4}{9}D + \frac{13}{27}D^2\right)x^2 = \frac{1}{9}x^3 - \frac{4}{9}x^2 + \frac{26}{27}x.$$

For $D^3 + 4D^2 + 3D = 0$ the roots are $0, -1, -3$ and the complementary function or solution of $P(D)y = 0$ would be $C_1 + C_2e^{-x} + C_3e^{-3x}$. Hence the solution of the equation $P(D)y = x^2$ is

$$y = C_1 + C_2e^{-x} + C_3e^{-3x} + \tfrac{1}{9}x^3 - \tfrac{4}{9}x^2 + \tfrac{26}{27}x.$$

It should be noted that in this example D is a factor of $P(D)$ and has been taken out before dividing; this shortens the work. Furthermore note that, in interpreting $1/D$ as integration, the constant may be omitted because any one value of I will do.

97. Next suppose that $X = Ce^{ax}$. Now $De^{ax} = ae^{ax}$, $D^ke^{ax} = a^ke^{ax}$,

and $$P(D)e^{ax} = P(a)e^{ax}; \quad \text{hence} \quad P(D)\left[\frac{C}{P(a)}e^{ax}\right] = Ce^{ax}.$$

But $$P(D)I = Ce^{ax}, \quad \text{and hence} \quad I = \frac{C}{P(a)}e^{ax} \tag{23}$$

is clearly a solution of the equation, provided a is not a root of $P(D) = 0$. If $P(a) = 0$, the division by $P(a)$ is impossible and the quest for I has to be directed more carefully. Let a be a root of multiplicity m so that $P(D) = (D-a)^mP_1(D)$. Then

$$P_1(D)(D-a)^mI = Ce^{ax}, \qquad (D-a)^mI = \frac{C}{P_1(a)}e^{ax},$$

and $$I = \frac{C}{P_1(a)}e^{ax}\int\cdots\int(dx)^m = \frac{Ce^{ax}x^m}{P_1(a)\,m!}. \tag{23'}$$

For in the integration the constants may be omitted. It follows that when $X = Ce^{ax}$, the solution I may be found *by direct substitution.*

Now if X broke up into the sum of terms $X = X_1 + X_2 + \cdots$ and if solutions $I_1, I_2, \cdots$ were determined for each of the equations $P(D)I_1 = X_1$, $P(D)I_2 = X_2, \cdots$, the solution I corresponding to X would be the sum $I_1 + I_2 + \cdots$. Thus it is seen that the above short methods apply to equations in which X is a sum of terms of the form Cx^m or Ce^{ax}.

As an example consider $(D^4 - 2D^2 + 1)y = e^x$. The roots are $1, 1, -1, -1$, and $a = 1$. Hence the solution for I is written as

$$(D+1)^2(D-1)^2I = e^x, \qquad (D-1)^2I = \tfrac{1}{4}e^x, \qquad I = \tfrac{1}{8}e^xx^2.$$

Then $$y = e^x(C_1 + C_2x) + e^{-x}(C_3 + C_4x) + \tfrac{1}{8}e^xx^2.$$

Again consider $(D^2 - 5D + 6)y = x + e^{mx}$. To find the I_1 corresponding to x, divide.

$$I_1 = \frac{1}{6 - 5D + D^2}x = \left(\frac{1}{6} + \frac{5}{36}D + \cdots\right)x = \frac{1}{6}x + \frac{5}{36}.$$

To find the I_2 corresponding to e^{mx}, substitute. There are three cases,

$$I_2 = \frac{1}{m^2 - 5m + 6}e^{mx}, \qquad I_2 = xe^{3x}, \qquad I_2 = -xe^{2x},$$

according as m is neither 2 nor 3, or is 3, or is 2. Hence for the complete solution,

$$y = C_1 e^{3x} + C_2 e^{2x} + \frac{1}{6}x + \frac{5}{36} + \frac{1}{m^2 - 5m + 6} e^{mx},$$

when m is neither 2 nor 3; but in these special cases the results are

$$y = C_1 e^{3x} + C_2 e^{2x} + \tfrac{1}{6}x + \tfrac{5}{36} - xe^{2x}, \qquad y = C_1 e^{3x} + C_2 e^{2x} + \tfrac{1}{6}x + \tfrac{5}{36} + xe^{3x}.$$

The next case to consider is where X is of the form $\cos \beta x$ or $\sin \beta x$. If these trigonometric functions be expressed in terms of exponentials, the solution may be conducted by the method above; and this is perhaps the best method when $\pm \beta i$ are roots of the equation $P(D) = 0$. It may be noted that this method would apply also to the case where X might be of the form $e^{ax} \cos \beta x$ or $e^{ax} \sin \beta x$. Instead of splitting the trigonometric functions into two exponentials, it is possible to combine two trigonometric functions into an exponential. Thus, consider the equations

$$P(D)y = e^{ax} \cos \beta x, \qquad P(D)y = e^{ax} \sin \beta x,$$

and
$$P(D)y = e^{ax}(\cos \beta x + i \sin \beta x) = e^{(\alpha + \beta i)x}. \tag{24}$$

The solution I of this last equation may be found and split into its real and imaginary parts, of which the real part is the solution of the equation involving the cosine, and the imaginary part the sine.

When X has the form $\cos \beta x$ or $\sin \beta x$ and $\pm \beta i$ are not roots of the equation $P(D) = 0$, there is a very short method of finding I. For

$$D^2 \cos \beta x = -\beta^2 \cos \beta x \quad \text{and} \quad D^2 \sin \beta x = -\beta^2 \sin \beta x.$$

Hence if $P(D)$ be written as $P_1(D^2) + DP_2(D^2)$ by collecting the even terms and the odd terms so that P_1 and P_2 are both even in D, the solution may be carried out symbolically as

$$I = \frac{1}{P(D)} \cos x = \frac{1}{P_1(D^2) + DP_2(D^2)} \cos x = \frac{1}{P_1(-\beta^2) + DP_2(-\beta^2)} \cos x,$$

or
$$I = \frac{P_1(-\beta^2) - DP_2(-\beta^2)}{[P_1(-\beta^2)]^2 + \beta^2 [P_2(-\beta^2)]^2} \cos x. \tag{25}$$

By this device of substitution and of rationalization as if D were a surd, the differentiation is transferred to the numerator and can be performed. This method of procedure may be justified directly, or it may be made to depend upon that of the paragraph above.

Consider the example $(D^2 + 1)y = \cos x$. Here $\beta i = i$ is a root of $D^2 + 1 = 0$. As an operator D^2 is equivalent to -1, and the rationalization method will not work. If the first solution be followed, the method of solution is

$$I = \frac{1}{D^2 + 1} \frac{e^{ix}}{2} + \frac{1}{D^2 + 1} \frac{e^{-ix}}{2} = \frac{1}{D - i} \frac{e^{ix}}{4i} - \frac{1}{D + i} \frac{e^{-ix}}{4i} = \frac{1}{4i}[xe^{ix} - xe^{-ix}] = \frac{1}{2} x \sin x.$$

If the second suggestion be followed, the solution may be found as follows:

$$(D^2+1)\,I=\cos x+i\sin x=e^{ix},\qquad I=\frac{1}{D^2+1}e^{ix}=\frac{xe^{ix}}{2i}.$$

Now $$I=\frac{x}{2i}(\cos x+i\sin x)=\frac{1}{2}x\sin x-\frac{1}{2}ix\cos x.$$

Hence $I=\frac{1}{2}x\sin x$ for $(D^2+1)\,I=\cos x$,

and $I=-\frac{1}{2}x\cos x$ for $(D^2+1)\,I=\sin x$.

The complete solution is $y=C_1\cos x+C_2\sin x+\frac{1}{2}x\sin x$,

and for $(D^2+1)\,y=\sin x$, $y=C_1\cos x+C_2\sin x-\frac{1}{2}x\cos x$.

As another example take $(D^2-3D+2)\,y=\cos x$. The roots are 1, 2, neither is equal to $\pm\beta i=\pm i$, and the method of rationalization is practicable. Then

$$I=\frac{1}{D^2-3D+2}\cos x=\frac{1}{1-3D}\cos x=\frac{1+3D}{10}\cos x=\frac{1}{10}(\cos x-3\sin x).$$

The complete solution is $y=C_1e^{-x}+C_2e^{-2x}+\frac{1}{10}(\cos x-3\sin x)$. The extreme simplicity of this substitution-rationalization method is noteworthy.

EXERCISES

1. By the general method solve the equations:

(α) $\dfrac{d^2y}{dx^2}+4\dfrac{dy}{dx}+3y=2e^{2x}$, ($\beta$) $\dfrac{d^3y}{dx^3}-3\dfrac{d^2y}{dx^2}+3\dfrac{dy}{dx}-y=e^x$,

(γ) $(D^2-4D+2)\,y=x$, (δ) $(D^3+D^2-4D-4)\,y=x$,

(ϵ) $(D^3+5D^2+6D)\,y=x$, (ζ) $(D^2+D+1)\,y=xe^x$,

(η) $(D^2+D+1)\,y=\sin 2x$, (θ) $(D^2-4)\,y=x+e^{2x}$,

(ι) $(D^2+3D+2)\,y=x+\cos x$, (κ) $(D^4-4D^2)\,y=1-\sin x$,

(λ) $(D^2+1)\,y=\cos x$, (μ) $(D^2+1)\,y=\sec x$, (ν) $(D^2+1)\,y=\tan x$.

2. By the rule write the solutions of these equations:

(α) $(D^2+3D+2)\,y=0$, (β) $(D^3+3D^2+D-5)\,y=0$,

(γ) $(D-1)^3y=0$, (δ) $(D^4+2D^2+1)\,y=0$,

(ϵ) $(D^3-3D^2+4)\,y=0$, (ζ) $(D^4-D^3-9D^2-11D-4)\,y=0$,

(η) $(D^3-6D^2+9D)\,y=0$, (θ) $(D^4-4D^3+8D^2-8D+4)\,y=0$,

(ι) $(D^5-2D^4+D^3)\,y=0$, (κ) $(D^3-D^2+D)\,y=0$,

(λ) $(D^4-1)^2y=0$, (μ) $(D^5-13D^3+26D^2+82D+104)\,y=0$.

3. By the short method solve (γ), (δ), (ϵ) of Ex. 1, and also:

(α) $(D^4-1)\,y=x^4$, (β) $(D^3-6D^2+11D-6)\,y=x$,

(γ) $(D^3+3D^2+2D)\,y=x^2$, (δ) $(D^3-3D^2-6D+8)\,y=x$,

(ϵ) $(D^3+8)\,y=x^4+2x+1$, (ζ) $(D^3-3D^2-D+3)\,y=x^2$,

(η) $(D^4-2D^3+D^2)\,y=x$, (θ) $(D^4+2D^3+3D^2+2D+1)\,y=1+x+x^2$,

(ι) $(D^3-1)\,y=x^2$, (κ) $(D^4-2D^3+D^2)\,y=x^3$.

4. By the short method solve (α), (β), (θ) of Ex. 1, and also:

(α) $(D^2-3D+2)\,y=e^x$, (β) $(D^4-D^3-3D^2+5D-2)\,y=e^{3x}$,

(γ) $(D^2-2D+1)\,y=e^x$, (δ) $(D^3-3D^2+4)\,y=e^{3x}$,

(ϵ) $(D^2+1)\,y=2e^x+x^3-x$, (ζ) $(D^3+1)\,y=3+e^{-x}+5e^{2x}$,

(η) $(D^4+2D^2+1)\,y=e^x+4$, (θ) $(D^3+3D^2+3D+1)\,y=2e^{-x}$,

(ι) $(D^2-2D)\,y=e^{2x}+1$, (κ) $(D^3+2D^2+D)\,y=e^{2x}+x^2+x$,

(λ) $(D^2-a^2)\,y=e^{ax}+e^{bx}$, ($\mu$) $(D^2-2aD+a^2)\,y=e^x+1$.

5. Solve by the short method (η), (ι), (κ) of Ex. 1, and also:

(α) $(D^2 - D - 2)y = \sin x$,
(β) $(D^2 + 2D + 1)y = 3e^{2x} - \cos x$,
(γ) $(D^2 + 4)y = x^2 + \cos x$,
(δ) $(D^3 + D^2 - D - 1)y = \cos 2x$,
(ϵ) $(D^2 + 1)^2 y = \cos x$,
(ζ) $(D^3 - D^2 + D - 1)y = \cos x$,
(η) $(D^2 - 5D + 6)y = \cos x - e^{2x}$,
(θ) $(D^3 - 2D^2 - 3D)y = 3x^2 + \sin x$,
(ι) $(D^2 - 1)^2 y = \sin x$,
(κ) $(D^2 + 3D + 2)y = e^{2x}\sin x$,
(λ) $(D^4 - 1)y = e^x \cos x$,
(μ) $(D^3 - 3D^2 + 4D - 2)y = e^x + \cos x$,
(ν) $(D^2 - 2D + 4)y = e^x \sin x$,
(o) $(D^2 + 4)y = \sin 3x + e^x + x^2$,
(π) $(D^6 + 1)y = \sin \frac{3}{2}x \sin \frac{1}{2}x$,
(ρ) $(D^3 + 1)y = e^{2x}\sin x + e^{\frac{x}{2}}\sin \dfrac{x\sqrt{3}}{2}$,
(σ) $(D^2 + 4)y = \sin^2 x$,
(τ) $(D^4 + 32D + 48)y = xe^{-2x} + e^{2x}\cos 2^{\frac{3}{2}}x$.

6. If X has the form $e^{ax}X_1$, show that $I = \dfrac{1}{P(D)}e^{ax}X_1 = e^{ax}\dfrac{1}{P(D+a)}X_1$. This enables the solution of equations where X_1 is a polynomial to be obtained by a short method; it also gives a way of treating equations where X is $e^{ax}\cos\beta x$ or $e^{ax}\sin\beta x$, but is not an improvement on (24); finally, combined with the second suggestion of (24), it covers the case where X is the product of a sine or cosine by a polynomial. Solve by this method, or partly by this method, (ζ) of Ex. 1; (κ), (λ), (ν), (ρ), (τ) of Ex. 5; and also

(α) $(D^2 - 2D + 1)y = x^2 e^{3x}$,
(β) $(D^3 + 3D^2 + 3D + 1)y = (2 - x^2)e^{-x}$,
(γ) $(D^2 + n^2)y = x^4 e^x$,
(δ) $(D^4 - 2D^3 - 3D^2 + 4D + 4)y = x^2 e^x$,
(ϵ) $(D^3 - 7D - 6)y = e^{2x}(1 + x)$,
(ζ) $(D - 1)^2 y = e^x + \cos x + x^2 e^x$,
(η) $(D - 1)^3 y = x - x^3 e^x$,
(θ) $(D^2 + 2)y = x^2 e^{3x} + e^x \cos 2x$,
(ι) $(D^3 - 1)y = xe^x + \cos^2 x$,
(κ) $(D^2 - 1)y = x\sin x + (1 + x^2)e^x$,
(λ) $(D^2 + 4)y = x\sin x$,
(μ) $(D^4 + 2D^2 + 1)y = x^2 \cos ax$,
(ν) $(D^2 + 4)y = (x\sin x)^2$,
(o) $(D^2 - 2D + 4)^2 y = xe^x \cos\sqrt{3}x$.

7. Show that the substitution $x = e^t$, Ex. 9, p. 152, changes equations of the type

$$x^n D^n y + a_1 x^{n-1} D^{n-1} y + \cdots + a_{n-1} x D y + a_n y = X(x) \tag{26}$$

into equations with constant coefficients: also that $ax + b = e^t$ would make a similar simplification for equations whose coefficients were powers of $ax + b$. Hence integrate:

(α) $(x^2D^2 - xD + 2)y = x\log x$,
(β) $(x^3D^3 - x^2D^2 + 2xD - 2)y = x^3 + 3x$,
(γ) $[(2x - 1)^3 D^3 + (2x - 1)D - 2]y = 0$,
(δ) $(x^2D^2 + 3xD + 1)y = (1 - x)^{-2}$,
(ϵ) $(x^3D^3 + xD - 1)y = x\log x$,
(ζ) $[(x + 1)^2 D^2 - 4(x + 1)D + 6]y = x$,
(η) $(x^2D^2 + 4xD + 2)y = e^x$,
(θ) $(x^3D^2 - 3x^2D + x)y = \log x \sin \log x + 1$,
(ι) $(x^4D^4 + 6x^3D^3 + 4x^2D^2 - 2xD - 4)y = x^2 + 2\cos\log x$.

8. If L be self-induction, R resistance, C capacity, i current, q charge upon the plates of a condenser, and $f(t)$ the electromotive force, then the differential equations for the circuit are

$$(\alpha)\quad \frac{d^2q}{dt^2} + \frac{R}{L}\frac{dq}{dt} + \frac{q}{LC} = \frac{1}{L}f(t), \qquad (\beta)\quad \frac{d^2i}{dt^2} + \frac{R}{L}\frac{di}{dt} + \frac{i}{LC} = \frac{1}{L}f'(t).$$

Solve (α) when $f(t) = e^{-at}\sin bt$ and (β) when $f(t) = \sin bt$. Reduce the trigonometric part of the particular solution to the form $K\sin(bt + \gamma)$. Show that if R is small and b is nearly equal to $1/\sqrt{LC}$, the amplitude K is large.

98. Simultaneous linear equations with constant coefficients. If there be given two (or in general n) linear equations with constant coefficients in two (or in general n) dependent variables and one independent variable t, the symbolic method of solution may still be used to advantage. Let the equations be

$$\begin{aligned}(a_0D^n + a_1D^{n-1} + \cdots + a_n)x + (b_0D^m + b_1D^{m-1} + \cdots + b_m)y = R(t),\\ (c_0D^p + c_1D^{p-1} + \cdots + c_p)x + (d_0D^q + d_1D^{q-1} + \cdots + d_q)y = S(t),\end{aligned} \tag{27}$$

when there are two variables and where D denotes differentiation by t. The equations may also be written more briefly as

$$P_1(D)x + Q_1(D)y = R \quad \text{and} \quad P_2(D)x + Q_2(D)y = S.$$

The ordinary algebraic process of solution for x and y may be employed because it depends only on such laws as are satisfied equally by the symbols D, $P_1(D)$, $Q_1(D)$, and so on.

Hence the solution for x and y is found by multiplying by the appropriate coefficients and adding the equations.

$$\begin{array}{rr|l} Q_2(D) & -P_2(D) & P_1(D)x + Q_1(D)y = R,\\ -Q_1(D) & P_1(D) & P_2(D)x + Q_2(D)y = S.\end{array}$$

Then
$$\begin{aligned}[P_1(D)Q_2(D) - P_2(D)Q_1(D)]x &= Q_2(D)R - Q_1(D)S,\\ [P_1(D)Q(D) - P_2(D)Q_1(D)]y &= P_1(D)S - P_2(D)R.\end{aligned} \tag{27'}$$

It will be noticed that the coefficients by which the equations are multiplied (written on the left) are so chosen as to make the coefficients of x and y in the solved form the same in sign as in other respects. It may also be noted that the order of P and Q in the symbolic products is immaterial. By expanding the operator $P_1(D)Q_2(D) - P_2(D)Q_1(D)$ a certain polynomial in D is obtained and by applying the operators to R and S as indicated certain functions of t are obtained. Each equation, whether in x or in y, is quite of the form that has been treated in §§ 95–97.

As an example consider the solution for x and y in the case of

$$2\frac{d^2x}{dt^2} - \frac{dy}{dt} - 4x = 2t, \qquad 2\frac{dx}{dt} + 4\frac{dy}{dt} - 3y = 0;$$

or
$$(2D^2 - 4)x - Dy = 2t, \qquad 2Dx + (4D - 3)y = 0.$$

Solve
$$\begin{array}{rr|l} 4D-3 & -2D & (2D^2 - 4)x - Dy = 2t\\ D & 2D^2 - 4 & 2Dx + (4D - 3)y = 0.\end{array}$$

Then
$$\begin{aligned}[(4D - 3)(2D^2 - 4) + 2D^2]x &= (4D - 3)2t,\\ [2D^2 + (2D^2 - 4)(4D - 3)]y &= -(2D)2t,\end{aligned}$$

or
$$4(2D^3 - D^2 - 4D + 3)x = 8 - 6t, \qquad 4(2D^3 - D^2 - 4D + 3)y = -4.$$

The roots of the polynomial in D are $1, 1, -1\frac{1}{2}$; and the particular solution I_x for x is $-\frac{1}{2}t$, and I_y for y is $-\frac{1}{3}$. Hence the solutions have the form

$$x = (C_1 + C_2t)e^t + C_3e^{-\frac{3}{2}t} - \tfrac{1}{2}t, \qquad y = (K_1 + K_2t)e^t + K_3e^{-\frac{3}{2}t} - \tfrac{1}{3}.$$

The arbitrary constants which are introduced into the solutions for x and y are not independent nor are they identical. *The solutions must be substituted into one of the equations to establish the necessary relations between the constants.* It will be noticed that in general the order of the equation in D for x and for y is the sum of the orders of the highest derivatives which occur in the two equations, — in this case, $3 = 2 + 1$. The order may be diminished by cancellations which occur in the formal algebraic solutions for x and y. In fact it is conceivable that the coefficient $P_1Q_2 - P_2Q_1$ of x and y in the solved equations should vanish and the solution become illusory. This case is of so little consequence in practice that it may be dismissed with the statement that the solution is then either impossible or indeterminate; that is, either there are no functions x and y of t which satisfy the two given differential equations, or there are an infinite number in each of which other things than the constants of integration are arbitrary.

To finish the example above and determine one set of arbitrary constants in terms of the other, substitute in the second differential equation. Then

$$2\left(C_1e^t + C_2e^t + C_2te^t - \tfrac{3}{2}C_3e^{-\frac{3}{2}t} - \tfrac{1}{2}\right) + 4\left(K_1e^t + K_2e^t + K_2te^t - \tfrac{3}{2}K_3e^{-\frac{3}{2}t}\right)$$
$$- 3\left(K_1e^t + K_2te^t + K_3e^{-\frac{3}{2}t} - \tfrac{1}{3}\right) = 0,$$

or $$e^t(2C_1 + 2C_2 + K_1 + K_2) + te^t(2C_2 + K_2) - 3e^{-\frac{3}{2}t}(C_3 + 3K_3) = 0.$$

As the terms e^t, te^t, $e^{-\frac{3}{2}t}$ are independent, the linear relation between them can hold only if each of the coefficients vanishes. Hence

$$C_3 + 3K_3 = 0, \qquad 2C_2 + K_2 = 0, \qquad 2C_1 + 2C_2 + K_1 + K_2 = 0,$$

and $$C_3 = -3K_3, \qquad 2C_2 = -K_2, \qquad 2C_1 = -K_1.$$

Hence $x = (C_1 + C_2t)e^t - 3K_3e^{-\frac{3}{2}t} - \frac{1}{2}t, \qquad y = -2(C_1 + C_2t)e^t + K_3e^{-\frac{3}{2}t} - \frac{1}{3}$

are the finished solutions, where C_1, C_2, K_3 are three arbitrary constants of integration and might equally well be denoted by C_1, C_2, C_3, or K_1, K_2, K_3.

99. One of the most important applications of the theory of simultaneous equations with constant coefficients is to *the theory of small vibrations about a state of equilibrium in a conservative* dynamical system.* If $q_1, q_2, \cdots, q_n$ are n coördinates (see Exs. 19–20, p. 112) which specify the position of the system measured relatively

* The potential energy V is defined as $-dV = dW = Q_1dq_1 + Q_2dq_2 + \cdots + Q_ndq_n$, where
$$Q_i = X_1\frac{\partial x_1}{\partial q_i} + Y_1\frac{\partial y_1}{\partial q_i} + Z_1\frac{\partial z_1}{\partial q_i} + \cdots + X_n\frac{\partial x_n}{\partial q_i} + Y_n\frac{\partial y_n}{\partial q_i} + Z_n\frac{\partial z_n}{\partial q_i}.$$
This is the immediate extension of Q_1 as given in Ex. 19, p. 112. Here dW denotes the differential of work and $dW = \Sigma \mathbf{F}_i \cdot d\mathbf{r}_i = \Sigma(X_idx_i + Y_idy_i + Z_idz_i)$. To find Q_i it is generally quickest to compute dW from this relation with dx_i, dy_i, dz_i expressed in terms of the differentials $dq_1, \cdots, dq_n$. The generalized forces Q_i are then the coefficients of dq_i. If there is to be a potential V, the differential dW must be exact. It is frequently easy to find V directly in terms of $q_1, \cdots, q_n$ rather than through the mediation of $Q_1, \cdots, Q_n$; when this is not so, it is usually better to leave the equations in the form $\dfrac{d}{dt}\dfrac{\partial T}{\partial \dot{q}_i} - \dfrac{\partial T}{\partial q_i} = Q_i$ rather than to introduce V and L.

to a position of stable equilibrium in which all the q's vanish, the development of the potential energy by Maclaurin's Formula gives

$$V(q_1, q_2, \cdots, q_n) = V_0 + V_1(q_1, q_2, \cdots, q_n) + V_2(q_1, q_2, \cdots, q_n) + \cdots,$$

where the first term is constant, the second is linear, and the third is quadratic, and where the supposition that the q's take on only small values, owing to the restriction to small vibrations, shows that each term is infinitesimal with respect to the preceding. Now the constant term may be neglected in any expression of potential energy. As the position when all the q's are 0 is assumed to be one of equilibrium, the forces

$$Q_1 = -\frac{\partial V}{\partial q_1}, \qquad Q_2 = -\frac{\partial V}{\partial q_2}, \quad \cdots, \quad Q_n = -\frac{\partial V}{\partial q_n}$$

must all vanish when the q's are 0. This shows that the coefficients, $(\partial V/\partial q_i)_0 = 0$, of the linear expression are all zero. Hence the first term in the expansion is the quadratic term, and relative to it the higher terms may be disregarded. As the position of equilibrium is stable, the system will tend to return to the position where all the q's are 0 when it is slightly displaced from that position. It follows that the quadratic expression must be definitely positive.

The kinetic energy is always a quadratic function of the velocities $\dot{q}_1, \dot{q}_2, \cdots, \dot{q}_n$ with coefficients which may be functions of the q's. If each coefficient be expanded by the Maclaurin Formula and only the first or constant term be retained, the kinetic energy becomes a quadratic function with constant coefficients. Hence the Lagrangian function (cf. § 160)

$$L = T - V = T(\dot{q}_1, \dot{q}_2, \cdots, \dot{q}_n) - V(q_1, q_2, \cdots, q_n),$$

when substituted in the formulas for the motion of the system, gives

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_1} - \frac{\partial L}{\partial q_1} = 0, \qquad \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_2} - \frac{\partial L}{\partial q_2} = 0, \quad \cdots, \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_n} - \frac{\partial L}{\partial q_n} = 0,$$

a set of equations of the second order with constant coefficients. The equations moreover involve the operator D only through its square, and the roots of the equation in D must be either real or pure imaginary. The pure imaginary roots introduce trigonometric functions in the solution and represent vibrations. If there were real roots, which would have to occur in pairs, the positive root would represent a term of exponential form which would increase indefinitely with the time, — a result which is at variance both with the assumption of stable equilibrium and with the fact that the energy of the system is constant.

When there is friction in the system, the forces of friction are supposed to vary with the velocities for small vibrations. In this case there exists a dissipative function $F(\dot{q}_1, \dot{q}_2, \cdots, \dot{q}_n)$ which is quadratic in the velocities and may be assumed to have constant coefficients. The equations of motion of the system then become

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_1} - \frac{\partial L}{\partial q_1} + \frac{\partial F}{\partial \dot{q}_1} = 0, \quad \cdots, \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_n} - \frac{\partial L}{\partial q_n} + \frac{\partial F}{\partial \dot{q}_n} = 0,$$

which are still linear with constant coefficients but involve first powers of the operator D. It is physically obvious that the roots of the equation in D must be negative if real, and must have their real parts negative if the roots are complex; for otherwise the energy of the motion would increase indefinitely with the time, whereas it is known to be steadily dissipating its initial energy. It may be added that if, in addition to the internal forces arising from the potential V and the

frictional forces arising from the dissipative function F, there are other forces impressed on the system, these forces would remain to be inserted upon the right-hand side of the equations of motion just given.

The fact that the equations for small vibrations lead to equations with constant coefficients by neglecting the higher powers of the variables gives the important physical theorem of the superposition of small vibrations. The theorem is: If with a certain set of initial conditions, a system executes a certain motion; and if with a different set of initial conditions taken at the same initial time, the system executes a second motion; then the system may execute the motion which consists of merely adding or superposing these motions at each instant of time; and in particular this combined motion will be that which the system would execute under initial conditions which are found by simply adding the corresponding values in the two sets of initial conditions. This theorem is of course a mere corollary of the linearity of the equations.

EXERCISES

1. Integrate the following systems of equations:

(α) $Dx - Dy + x = \cos t,\qquad D^2x - Dy + 3x - y = e^{2t},$

(β) $3Dx + 3x + 2y = e^t,\qquad 4x - 3Dy + 3y = 3t,$

(γ) $D^2x - 3x - 4y = 0,\qquad D^2y + x + y = 0,$

(δ) $\dfrac{dx}{y - 7x} = \dfrac{-dy}{2x + 5y} = dt,\qquad$ (ϵ) $-dt = \dfrac{dx}{3x + 4y} = \dfrac{dy}{2x + 5y},$

(ζ) $tDx + 2(x - y) = 1,\qquad tDy + x + 5y = t,$

(η) $Dx = ny - mz,\qquad Dy = lz - nx,\qquad Dz = mx - ly,$

(θ) $D^2x - 3x - 4y + 3 = 0,\qquad D^2y + x - 8y + 5 = 0,$

(ι) $D^4x - 4D^3y + 4D^2x - x = 0,\qquad D^4y - 4D^3x + 4D^2y - y = 0.$

2. A particle vibrates without friction upon the inner surface of an ellipsoid. Discuss the motion. Take the ellipsoid as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{(z - c)^2}{c^2} = 1; \quad \text{then} \quad x = C\sin\left(\frac{\sqrt{cg}}{a}t + C_1\right), \quad y = K\sin\left(\frac{\sqrt{cg}}{b}t + K_1\right).$$

3. Same as Ex. 2 when friction varies with the velocity.

4. Two heavy particles of equal mass are attached to a light string, one at the middle, one at one end, and are suspended by attaching the other end of the string to a fixed point. If the particles are slightly displaced and the oscillations take place without friction in a vertical plane containing the fixed point, discuss the motion.

5. If there be given two electric circuits without capacity, the equations are

$$L_1\frac{di_1}{dt} + M\frac{di_2}{dt} + R_1i_1 = E_1, \qquad L_2\frac{di_2}{dt} + M\frac{di_1}{dt} + R_2i_2 = E_2,$$

where i_1, i_2 are the currents in the circuits, L_1, L_2 are the coefficients of self-induction, R_1, R_2 are the resistances, and M is the coefficient of mutual induction. (α) Integrate the equations when the impressed electromotive forces E_1, E_2 are zero in both circuits. (β) Also when $E_2 = 0$ but $E_1 = \sin pt$ is a periodic force. (γ) Discuss the cases of loose coupling, that is, where M^2/L_1L_2 is small; and the case of close coupling, that is, where M^2/L_1L_2 is nearly unity. What values for p are especially noteworthy when the damping is small?

6. If the two circuits of Ex. 5 have capacities C_1, C_2 and if q_1, q_2 are the charges on the condensers so that $i_1 = dq_1/dt$, $i_2 = dq_2/dt$ are the currents, the equations are

$$L_1 \frac{d^2q_1}{dt^2} + M \frac{d^2q_2}{dt^2} + R_1 \frac{dq_1}{dt} + \frac{q_1}{C_1} = E_1, \qquad L_2 \frac{d^2q_2}{dt^2} + M \frac{d^2q_1}{dt^2} + R_2 \frac{dq_2}{dt} + \frac{q_2}{C_2} = E_2.$$

Integrate when the resistances are negligible and $E_1 = E_2 = 0$. If $T_1 = 2\pi\sqrt{C_1L_1}$ and $T_2 = 2\pi\sqrt{C_2L_2}$ are the periods of the individual separate circuits and $\Theta = 2\pi M\sqrt{C_1C_2}$, and if $T_1 = T_2$, show that $\sqrt{T^2 + \Theta^2}$ and $\sqrt{T^2 - \Theta^2}$ are the independent periods in the coupled circuits.

7. A uniform beam of weight 6 lb. and length 2 ft. is placed orthogonally across a rough horizontal cylinder 1 ft. in diameter. To each end of the beam is suspended a weight of 1 lb. upon a string 1 ft. long. Solve the motion produced by giving one of the weights a slight horizontal velocity. Note that in finding the kinetic energy of the beam, the beam may be considered as rotating about its middle point (§ 39).

CHAPTER IX

ADDITIONAL TYPES OF ORDINARY EQUATIONS

100. Equations of the first order and higher degree. The *degree* of a differential equation is defined as the degree of the derivative of highest order which enters in the equation. In the case of the equation $\Psi(x, y, y') = 0$ of the first order, the degree will be the degree of the equation in y'. From the idea of the lineal element (§ 85) it appears that if the degree of Ψ in y' is n, there will be n lineal elements through each point (x, y). Hence it is seen that there are n curves, which are compounded of these elements, passing through each point. It may be pointed out that equations such as $y' = x\sqrt{1 + y^2}$, which are apparently of the first degree in y', are really of higher degree if the multiple value of the functions, such as $\sqrt{1 + y^2}$, which enter in the equation, is taken into consideration; the equation above is replaceable by $y'^2 = x^2 + x^2y^2$, which is of the second degree and without any multiple valued function.*

First suppose that *the differential equation*

$$\Psi(x, y, y') = [y' - \psi_1(x, y)] \times [y' - \psi_2(x, y)] \cdots = 0 \tag{1}$$

may be solved for y'. It then becomes equivalent to the set

$$y' - \psi_1(x, y) = 0, \qquad y' - \psi_2(x, y) = 0, \cdots \tag{1'}$$

of equations each of the first order, and each of these may be treated by the methods of Chap. VIII. Thus a set of integrals †

$$F_1(x, y, C) = 0, \qquad F_2(x, y, C) = 0, \cdots \tag{2}$$

may be obtained, and the product of these separate integrals

$$F(x, y, C) = F_1(x, y, C) \cdot F_2(x, y, C) \cdots = 0 \tag{2'}$$

is the complete solution of the original equation. Geometrically speaking, each integral $F_i(x, y, C) = 0$ represents a family of curves and the product represents all the families simultaneously.

* It is therefore apparent that the idea of degree as applied in practice is somewhat indefinite.

† The same constant C or any desired function of C may be used in the different solutions because C is an arbitrary constant and no specialization is introduced by its repeated use in this way.

As an example consider $y'^2 + 2\,y'y\cot x = y^2$. Solve.

$$y'^2 + 2\,y'y\cot x + y^2\cot^2 x = y^2(1 + \cot^2 x) = y^2\csc^2 x,$$

and
$$(y' + y\cot x - y\csc x)(y' + y\cot x + y\csc x) = 0.$$

These equations both come under the type of variables separable. Integrate

$$\frac{dy}{y} = \frac{1 - \cos x}{\sin x}\,dx = -\frac{d\cos x}{1 + \cos x}, \qquad y\,(1 + \cos x) = C,$$

and
$$\frac{dy}{y} = -\frac{1 + \cos x}{\sin x}\,dx = \frac{d\cos x}{1 - \cos x}, \qquad y\,(1 - \cos x) = C.$$

Hence
$$[y\,(1 + \cos x) + C][y\,(1 - \cos x) + C] = 0$$

is the solution. It may be put in a different form by multiplying out. Then

$$y^2\sin^2 x + 2\,Cy + C^2 = 0.$$

If the equation cannot be solved for y' or if the equations resulting from the solution cannot be integrated, this first method fails. In that case *it may be possible to solve for y or for x* and treat the equation by differentiation. Let $y' = p$. Then if

$$y = f(x, p), \qquad \frac{dy}{dx} = p = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial p}\frac{dp}{dx}. \tag{3}$$

The equation thus found by differentiation is a differential equation of the first order in dp/dx and it may be solved by the methods of Chap. VIII to find $F(p, x, C) = 0$. The two equations

$$y = f(x, p) \quad \text{and} \quad F(p, x, C) = 0 \tag{3'}$$

may be regarded as defining x and y parametrically in terms of p, or p may be eliminated between them to determine the solution in the form $\Omega(x, y, C) = 0$ if this is more convenient. If the given differential equation had been solved for x, then

$$x = f(y, p) \quad \text{and} \quad \frac{dx}{dy} = \frac{1}{p} = \frac{\partial f}{\partial y} + \frac{\partial f}{\partial p}\frac{dp}{dy}. \tag{4}$$

The resulting equation on the right is an equation of the first order in dp/dy and may be treated in the same way.

As an example take $xp^2 - 2\,yp + ax = 0$ and solve for y. Then

$$2\,y = xp + \frac{ax}{p}, \qquad 2\frac{dy}{dx} = 2\,p = p + x\frac{dp}{dx} - \frac{ax}{p^2}\frac{dp}{dx} + \frac{a}{p},$$

or
$$\frac{x}{p}\left[p - \frac{a}{p}\right]\frac{dp}{dx} + \left(\frac{a}{p} - p\right) = 0, \quad \text{or} \quad x\,dp - p\,dx = 0.$$

The solution of this equation is $x = Cp$. The solution of the given equation is

$$2\,y = xp + \frac{ax}{p}, \qquad x = Cp$$

when expressed parametrically in terms of p. If p be eliminated, then

$$2\,y = \frac{x^2}{C} + aC \qquad \text{parabolas.}$$

As another example take $p^2y + 2px = y$ and solve for x. Then

$$2x = y\left(\frac{1}{p} - p\right), \qquad 2\frac{dx}{dy} = \frac{2}{p} = \frac{1}{p} - p + y\left(-\frac{1}{p^2} - 1\right)\frac{dp}{dy},$$

or
$$\frac{1}{p} + p + y\left(\frac{1}{p^2} + 1\right)\frac{dp}{dy} = 0, \quad \text{or} \quad ydp + pdy = 0.$$

The solution of this is $py = C$ and the solution of the given equation is

$$2x = y\left(\frac{1}{p} - p\right), \qquad py = C, \quad \text{or} \quad y^2 = 2Cx + C^2.$$

Two special types of equation may be mentioned in addition, although their method of solution is a mere corollary of the methods already given in general. They are the equation *homogeneous* in (x, y) and *Clairaut's* equation. The general form of the homogeneous equation is $\Psi(p, y/x) = 0$. This equation may be solved as

$$p = \psi\left(\frac{y}{x}\right) \qquad \text{or as} \quad \frac{y}{x} = f(p), \qquad y = xf(p); \tag{5}$$

and in the first case is treated by the methods of Chap. VIII, and in the second by the methods of this article. Which method is chosen rests with the solver. The Clairaut type of equation is

$$y = px + f(p) \tag{6}$$

and comes directly under the methods of this article. It is especially noteworthy, however, that on differentiating with respect to x the resulting equation is

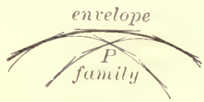
$$[x + f'(p)]\frac{dp}{dx} = 0 \quad \text{or} \quad \frac{dp}{dx} = 0. \tag{6'}$$

Hence the solution for p is $p = C$, and thus $y = Cx + f(C)$ is the solution for the Clairaut equation and represents a family of straight lines. The rule is merely to substitute C in place of p. This type occurs very frequently in geometric applications either directly or in a disguised form requiring a preliminary change of variable.

101. To this point the only solution of the differential equation $\Psi(x, y, p) = 0$ which has been considered is the *general solution* $F(x, y, C) = 0$ containing an arbitrary constant. If a special value, say 2, is given to C, the solution $F(x, y, 2) = 0$ is called a *particular solution*. It may happen that the arbitrary constant C enters into the expression $F(x, y, C) = 0$ in such a way that when C becomes positively infinite (or negatively infinite) the curve $F(x, y, C) = 0$ approaches a definite limiting position which is a solution of the differential equation; such solutions are called *infinite solutions*. In addition to these types of solution which naturally group themselves in connection with the general solution, there is often a solution of a different kind which is

known as the *singular solution.* There are several different definitions for the singular solution. That which will be adopted here is: *A singular solution is the envelope of the family of curves defined by the general solution.*

The consideration of the lineal elements (§ 85) will show how it is that the envelope (§ 65) of the family of particular solutions which constitute the general solution is itself a solution of the equation. For consider the figure, which represents the particular solutions broken up into their lineal elements. Note that the envelope is made up of those lineal elements, one taken from each particular solution, which are at the points of contact of the envelope with the curves of the family. It is seen that the envelope is a curve all of whose lineal elements satisfy the equation $\Psi(x, y, p) = 0$ for the reason that they lie upon solutions of the equation. Now any curve whose lineal elements satisfy the equation is by definition a solution of the equation; and so the envelope must be a solution. It might conceivably happen that the family $F(x, y, C) = 0$ was so constituted as to envelope one of its own curves. In that case that curve would be both a particular and a singular solution.



If the general solution $F(x, y, C) = 0$ of a given differential equation is known, the singular solution may be found according to the rule for finding envelopes (§ 65) by eliminating C from

$$F(x, y, C) = 0 \quad \text{and} \quad \frac{\partial}{\partial C} F(x, y, C) = 0. \tag{7}$$

It should be borne in mind that in the eliminant of these two equations there may occur some factors which do not represent envelopes and which must be discarded from the singular solution. If only the singular solution is desired and the general solution is not known, this method is inconvenient. In the case of Clairaut's equation, however, where the solution is known, it gives the result immediately as that obtained by eliminating C from the two equations

$$y = Cx + f(C) \quad \text{and} \quad 0 = x + f'(C). \tag{8}$$

It may be noted that as $p = C$, the second of the equations is merely the factor $x + f'(p) = 0$ discarded from (6′). The singular solution may therefore be found by eliminating p between the given Clairaut equation and the discarded factor $x + f'(p) = 0$.

A reëxamination of the figure will suggest a means of finding the singular solution without integrating the given equation. For it is seen that when two neighboring curves of the family intersect in a point P

near the envelope, then through this point there are two lineal elements which satisfy the differential equation. These two lineal elements have nearly the same direction, and indeed the nearer the two neighboring curves are to each other the nearer will their intersection lie to the envelope and the nearer will the two lineal elements approach coincidence with each other and with the element upon the envelope at the point of contact. Hence for all points (x, y) on the envelope the equation $\Psi(x, y, p) = 0$ of the lineal elements must have *double roots for p*. Now if an equation has double roots, the derivative of the equation must have a root. Hence the requirement that the two equations

$$\psi(x, y, p) = 0 \quad \text{and} \quad \frac{\partial}{\partial p}\psi(x, y, p) = 0 \tag{9}$$

have a common solution for p will insure that the first has a double root for p; and the points (x, y) which satisfy these equations simultaneously must surely include all the points of the envelope. The rule for finding the singular solution is therefore: *Eliminate p from the given differential equation and its derivative with respect to p*, that is, from (9). The result should be tested.

If the equation $xp^2 - 2yp + ax = 0$ treated above be tried for a singular solution, the elimination of p is required between the two equations

$$xp^2 - 2yp + ax = 0 \quad \text{and} \quad xp - y = 0.$$

The result is $y^2 = ax^2$, which gives a pair of lines through the origin. The substitution of $y = \pm\sqrt{a}x$ and $p = \pm\sqrt{a}$ in the given equation shows at once that $y^2 = ax^2$ satisfies the equation. Thus $y^2 = ax^2$ is a singular solution. The same result is found by finding the envelope of the general solution given above. It is clear that in this case the singular solution is not a particular solution, as the particular solutions are parabolas.

If the elimination had been carried on by Sylvester's method, then

$$\begin{vmatrix} 0 & x & -y \\ x & -2y & a \\ x & -y & 0 \end{vmatrix} = -x(y^2 - ax^2) = 0;$$

and the eliminant is the product of two factors $x = 0$ and $y^2 - ax^2 = 0$, of which the second is that just found and the first is the y-axis. As the slope of the y-axis is infinite, the substitution in the equation is hardly legitimate, and the equation can hardly be said to be satisfied. The occurrence of these extraneous factors in the eliminant is the real reason for the necessity of testing the result to see if it actually represents a singular solution. These extraneous factors may represent a great variety of conditions. Thus in the case of the equation $p^2 + 2yp\cot x = y^2$ previously treated, the elimination gives $y^2\csc^2 x = 0$, and as $\csc x$ cannot vanish, the result reduces to $y^2 = 0$, or the x-axis. As the slope along the x-axis is 0 and y is 0, the equation is clearly satisfied. Yet the line $y = 0$ is *not* the envelope of the general solution; for the curves of the family touch the line only at the points $n\pi$. It is a particular solution and corresponds to $C = 0$. There is no singular solution.

Many authors use a great deal of time and space discussing just what may and what may not occur among the extraneous loci and how many times it may occur. The result is a considerable number of statements which in their details are either grossly incomplete or glaringly false or both (cf. §§ 65–67). The rules here given for finding singular solutions should not be regarded in any other light than as leading to some expressions which are to be examined, the best way one can, to find out whether or not they are singular solutions. One curve which may appear in the elimination of p and which deserves a note is the *tac-locus* or locus of points of tangency of the particular solutions with each other. Thus in the system of circles $(x - C)^2 + y^2 = r^2$ there may be found two which are tangent to each other at any assigned point of the x-axis. This tangency represents two coincident lineal elements and hence may be expected to occur in the elimination of p between the differential equation of the family and its derivative with respect to p; but not in the eliminant from (7).

EXERCISES

1. Integrate the following equations by solving for $p = y'$:

(α) $p^2 - 6p + 5 = 0$, (β) $p^3 - (2x + y^2)p^2 + (x^2 - y^2 + 2xy^2)p - (x^2 - y^2)y^2 = 0$,
(γ) $xp^2 - 2yp - x = 0$, (δ) $p^3(x + 2y) + 3p^2(x + y) + p(y + 2x) = 0$,
(ϵ) $y^2 + p^2 = 1$, (ζ) $p^2 - ax^3 = 0$, (η) $p = (a - x)\sqrt{1 + p^2}$.

2. Integrate the following equations by solving for y or x:

(α) $4xp^2 + 2xp - y = 0$, (β) $y = -xp + x^4p^2$, (γ) $p + 2xy - x^2 - y^2 = 0$,
(δ) $2px - y + \log p = 0$, (ϵ) $x - yp = ap^2$, (ζ) $y = x + a\tan^{-1}p$,
(η) $x = y + a\log p$, (θ) $x + py(2p^2 + 3) = 0$, (ι) $a^2yp^2 - 2xp + y = 0$,
(κ) $p^3 - 4xyp + 8y^2 = 0$, (λ) $x = p + \log p$, (μ) $p^2(x^2 + 2ax) = a^2$.

3. Integrate these equations [substitutions suggested in (ι) and (κ)]:

(α) $xy^2(p^2 + 2) = 2py^3 + x^3$, ($\beta$) $(nx + py)^2 = (1 + p^2)(y^2 + nx^2)$,
(γ) $y^2 + xyp - x^2p^2 = 0$, (δ) $y = yp^2 + 2px$,
(ϵ) $y = px + \sin^{-1}p$, (ζ) $y = p(x - b) + a/p$,
(η) $y = px + p(1 - p^2)$, (θ) $y^2 - 2pxy - 1 = p^2(1 - x^2)$,
(ι) $4e^{2y}p^2 + 2xp - 1 = 0$, $z = e^{2y}$, (κ) $y = 2px + y^2p^3$, $y^2 = z$,
(λ) $4e^{2y}p^2 + 2e^{2x}p - e^x = 0$, ($\mu$) $x^2(y - px) = yp^2$.

4. Treat these equations by the p method (9) to find the singular solutions. Also solve and treat by the C method (7). Sketch the family of solutions and examine the significance of the extraneous factors as well as that of the factor which gives the singular solution:

(α) $p^2y + p(x - y) - x = 0$, (β) $p^2y^2\cos^2\alpha - 2pxy\sin^2\alpha + y^2 - x^2\sin^2\alpha = 0$,
(γ) $4xp^2 = (3x - a)^2$, (δ) $yp^2x(x - a)(x - b) = [3x^2 - 2x(a + b) + ab]^2$,
(ϵ) $p^2 + xp - y = 0$, (ζ) $8a(1 + p)^3 = 27(x + y)(1 - p)^3$,
(η) $x^3p^2 + x^2yp + a^3 = 0$, (θ) $y(3 - 4y)^2p^2 = 4(1 - y)$.

5. Examine sundry of the equations of Exs. 1, 2, 3, for singular solutions.

6. Show that the solution of $y = x\phi(p) + f(p)$ is given parametrically by the given equation and the solution of the linear equation:

$$\frac{dx}{dp} + x\frac{\phi'(p)}{\phi(p) - p} = \frac{f'(p)}{p - \phi(p)}. \qquad \text{Solve} \quad (\alpha)\ y = mxp + n(1 + p^3)^{\frac{1}{3}},$$

(β) $y = x(p + a\sqrt{1 + p^2})$, ($\gamma$) $x = yp + ap^2$, (δ) $y = (1 + p)x + p^2$.

7. As any straight line is $y = mx + b$, any family of lines may be represented as $y = mx + f(m)$ or by the Clairaut equation $y = px + f(p)$. Show that the orthogonal trajectories of any family of lines leads to an equation of the type of Ex. 6. The same is true of the trajectories at any constant angle. Express the equations of the following systems of lines in the Clairaut form, write the equations of the orthogonal trajectories, and integrate:

(α) tangents to $x^2 + y^2 = 1$, (β) tangents to $y^2 = 2ax$,
(γ) tangents to $y^2 = x^3$, (δ) normals to $y^2 = 2ax$,
(ϵ) normals to $y^2 = x^3$, (ζ) normals to $b^2x^2 + a^2y^2 = a^2b^2$.

8. The *evolute* of a given curve is the locus of the center of curvature of the curve, or, what amounts to the same thing, it is the envelope of the normals of the given curve. If the Clairaut equation of the normals is known, the evolute may be obtained as its singular solution. Thus find the evolutes of

(α) $y^2 = 4ax$, (β) $2xy = a^2$, (γ) $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$,
(δ) $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, ($\epsilon$) $y^2 = \dfrac{x^3}{2a - x}$, ($\zeta$) $y = \frac{1}{2}(e^x + e^{-x})$.

9. The *involutes* of a given curve are the curves which cut the tangents of the given curve orthogonally, or, what amounts to the same thing, they are the curves which have the given curve as the locus of their centers of curvature. Find the involutes of

(α) $x^2 + y^2 = a^2$, (β) $y^2 = 2mx$, (γ) $y = a\cosh(x/a)$.

10. As any curve is the envelope of its tangents, it follows that when the curve is described by a property of its tangents the curve may be regarded as the singular solution of the Clairaut equation of its tangent lines. Determine thus what curves have these properties:

(α) length of the tangent intercepted between the axes is l,
(β) sum of the intercepts of the tangent on the axes is c,
(γ) area between the tangent and axes is the constant k^2,
(δ) product of perpendiculars from two fixed points to tangent is k^2,
(ϵ) product of ordinates from two points of x-axis to tangent is k^2.

11. From the relation $\dfrac{dF}{dn} = \mu\sqrt{M^2 + N^2}$ of Proposition 3, p. 212, show that as the curve $F = C$ is moving tangentially to itself along its envelope, the singular solution of $Mdx + Ndy = 0$ may be expected to be found in the equation $1/\mu = 0$; also the infinite solutions. Discuss the equation $1/\mu = 0$ in the following cases:

(α) $\sqrt{1 - y^2}\,dx = \sqrt{1 - x^2}\,dy$, ($\beta$) $xdx + ydy = \sqrt{x^2 + y^2 - a^2}\,dy$.

102. Equations of higher order. In the treatment of special problems (§ 82) it was seen that the substitutions

$$\frac{dy}{dx} = p, \qquad \frac{d^2y}{dx^2} = \frac{dp}{dx} \quad \text{or} \quad \frac{d^2y}{dx^2} = p\frac{dp}{dy} \tag{10}$$

rendered the differential equations integrable by reducing them to integrable equations of the first order. These substitutions or others like them are useful in treating certain cases of the differential equation

$\Psi(x, y, y', y'', \cdots, y^{(n)}) = 0$ of the nth order, namely, when one of the variables and perhaps some of the derivatives of lowest order do not occur in the equation.

In case
$$\Psi\left(x, \frac{d^i y}{dx^i}, \frac{d^{i+1} y}{dx^{i+1}}, \cdots, \frac{d^n y}{dx^n}\right) = 0, \tag{11}$$

y and the first $i-1$ derivatives being absent, substitute

$$\frac{d^i y}{dx^i} = q \quad \text{so that} \quad \Psi\left(x, q, \frac{dq}{dx}, \cdots, \frac{d^{n-i} q}{dx^{n-i}}\right) = 0. \tag{11'}$$

The original equation is therefore replaced by one of lower order. If the integral of this be $F(x, q) = 0$, which will of course contain $n - i$ arbitrary constants, the solution for q gives

$$q = f(x) \quad \text{and} \quad y = \int \cdots \int f(x)\,(dx)^i. \tag{12}$$

The solution has therefore been accomplished. If it were more convenient to solve $F(x, q) = 0$ for $x = \phi(q)$, the integration would be

$$y = \int \cdots \int q\,(dx)^i = \int \cdots \int q\,[\phi'(q)\,dq]^i; \tag{12'}$$

and this equation with $x = \phi(q)$ would give a parametric expression for the integral of the differential equation.

In case
$$\Psi\left(y, \frac{dy}{dx}, \frac{d^2 y}{dx^2}, \cdots, \frac{d^n y}{dx^n}\right) = 0, \tag{13}$$

x being absent, substitute p and regard p as a function of y. Then

$$\frac{dy}{dx} = p, \qquad \frac{d^2 y}{dx^2} = p\frac{dp}{dy}, \qquad \frac{d^3 y}{dx^3} = p\frac{d}{dy}\left(p\frac{dp}{dy}\right), \cdots$$
and
$$\Psi_1\left(y, p, \frac{dp}{dy}, \cdots, \frac{d^{n-1} p}{dy^{n-1}}\right) = 0. \tag{13'}$$

In this way the order of the equation is lowered by unity. If this equation can be integrated as $F(y, p) = 0$, the last step in the solution may be obtained either directly or parametrically as

$$p = f(y), \qquad \int \frac{dy}{f(y)} = x \tag{14}$$

or
$$y = \phi(p), \qquad x = \int \frac{dy}{p} = \int \frac{\phi'(p)\,dp}{p}. \tag{14'}$$

It is no particular simplification in this case to have some of the lower derivatives of y absent from $\Psi = 0$, because in general the lower derivatives of p will none the less be introduced by the substitution that is made.

As an example consider $\left(x\frac{d^3y}{dx^3} - \frac{d^2y}{dx^2}\right)^2 = \left(\frac{d^3y}{dx^3}\right)^2 + 1,$

which is
$$\left(x\frac{dq}{dx} - q\right)^2 = \left(\frac{dq}{dx}\right)^2 + 1 \quad \text{if} \quad q = \frac{d^2y}{dx^2}.$$

Then
$$q = x\frac{dq}{dx} \pm \sqrt{\left(\frac{dq}{dx}\right)^2 + 1} \quad \text{and} \quad q = C_1x \pm \sqrt{C_1^2 + 1};$$

for the equation is a Clairaut type. Hence, finally,

$$y = \iint \left[C_1x \pm \sqrt{C_1^2 + 1}\right](dx)^2 = \tfrac{1}{6}\, C_1x^3 \pm \tfrac{1}{2}\, x^2 \sqrt{C_1^2 + 1} + C_2x + C_3.$$

As another example consider $y'' - y'^2 = y^2 \log y$. This becomes

$$p\frac{dp}{dy} - p^2 = y^2 \log y \quad \text{or} \quad \frac{d(p^2)}{dy} - 2p^2 = 2y^2 \log y.$$

The equation is linear in p^2 and has the integrating factor e^{-2y}.

$$\frac{1}{2}p^2e^{-2y} = \int y^2e^{-2y} \log y\,dy, \qquad \frac{1}{\sqrt{2}}\,p = \left[e^{2y}\int y^2e^{-2y}\log y\,dy\right]^{\frac{1}{2}},$$

and
$$\int \frac{dy}{\left[e^{2y}\int y^2e^{-2y}\log y\,dy\right]^{\frac{1}{2}}} = \sqrt{2}\,x.$$

The integration is therefore reduced to quadratures and becomes a problem in ordinary integration.

If an equation is *homogeneous with respect to y and its derivatives*, that is, if the equation is multiplied by a power of k when y is replaced by ky, the order of the equation may be lowered by the substitution $y = e^z$ and by taking z' as the new variable. If the equation is *homogeneous with respect to x and dx*, that is, if the equation is multiplied by a power of k when x is replaced by kx, the order of the equation may be reduced by the substitution $x = e^t$. The work may be simplified (Ex. 9, p. 152) by the use of

$$D_x^n y = e^{-nt}D_t(D_t - 1)\cdots(D_t - n + 1)\,y. \tag{15}$$

If the equation is *homogeneous with respect to x and y and the differentials dx, dy, d^2y, $\cdots$*, the order may be lowered by the substitution $x = e^t$, $y = e^tz$, where it may be recalled that

$$\begin{aligned} D_x^n y &= e^{-nt}D_t(D_t - 1)\cdots(D_t - n + 1)\,y \\ &= e^{-(n-1)t}(D_t + 1)\,D_t\cdots(D_t - n + 2)\,z. \end{aligned} \tag{15'}$$

Finally, if the equation is *homogeneous with respect to x considered of dimensions 1, and y considered of dimensions m*, that is, if the equation is multiplied by a power of k when kx replaces x and k^my replaces y, the substitution $x = e^t$, $y = e^{mt}z$ will lower the degree of the equation. It may be recalled that

$$D_x^n y = e^{(m-n)t}(D_t + m)(D_t + m - 1)\cdots(D_t + m - n + 1)\,z. \tag{15''}$$

Consider $xyy'' - xy'^2 = yy' + bxy'^2/\sqrt{a^2 - x^2}$. If in this equation y be replaced by ky so that y' and y'' are also replaced by ky' and ky'', it appears that the equation is merely multiplied by k^2 and is therefore homogeneous of the first sort mentioned. Substitute

$$y = e^z, \qquad y' = e^z z', \qquad y'' = e^z(z'' + z'^2).$$

Then e^{2z} will cancel from the whole equation, leaving merely

$$xz'' = z' + bxz'^2/\sqrt{a^2 - x^2} \quad \text{or} \quad \frac{xdz'}{z'^2} - \frac{1}{z'}dx = \frac{bxdx}{\sqrt{a^2 - x^2}}.$$

The equation in the first form is Bernoulli; in the second form, exact. Then

$$\frac{x}{z'} = b\sqrt{a^2 - x^2} + C \quad \text{and} \quad dz = \frac{xdx}{b\sqrt{a^2 - x^2} + C}.$$

The variables are separated for the last integration which will determine $z = \log y$ as a function of x.

Again consider $x^4 \dfrac{d^2y}{dx^2} = (x^3 + 2xy)\dfrac{dy}{dx} - 4y^2$. If x be replaced by kx and y by k^2y so that y' is replaced by ky' and y'' remains unchanged, the equation is multiplied by k^4 and hence comes under the fourth type mentioned above. Substitute

$$x = e^t, \qquad y = e^{2t}z, \qquad D_x y = e^t(D_t + 2)z, \qquad D_x^2 y = (D_t + 2)(D_t + 1)z.$$

Then e^{4t} will cancel and leave $z'' + 2(1 - z)z' = 0$, if accents denote differentiation with respect to t. This equation lacks the independent variable t and is reduced by the substitution $z'' = z'dz'/dz$. Then

$$\frac{dz'}{dz} + 2(1 - z) = 0, \qquad z' = \frac{dz}{dt} = (1 - z)^2 + C, \qquad \frac{dz}{(1 - z^2) + C} = dt.$$

There remains only to perform the quadrature and replace z and t by x and y.

103. If the equation may be obtained by differentiation, as

$$\Psi\left(x, y, \frac{dy}{dx}, \cdots, \frac{d^n y}{dx^n}\right) = \frac{d\Omega}{dx} = \frac{\partial\Omega}{\partial x} + \frac{\partial\Omega}{\partial y}y' + \cdots + \frac{\partial\Omega}{\partial y^{(n-1)}}y^{(n)}, \quad (16)$$

it is called an *exact equation*, and $\Omega(x, y, y', \cdots, y^{(n-1)}) = C$ is an integral of $\Psi = 0$. Thus in case the equation is exact, the order may be lowered by unity. It may be noted that unless the degree of the nth derivative is 1 the equation cannot be exact. Consider

$$\Psi(x, y, y', \cdots, y^{(n)}) = \phi_1 y^{(n)} + \phi_2,$$

where the coefficient of $y^{(n)}$ is collected into ϕ_1. Now integrate ϕ_1, partially regarding only $y^{(n-1)}$ as variable so that

$$\int \phi_1 dy^{(n-1)} = \Omega_1, \qquad \frac{d}{dx}\Omega_1 = \frac{\partial\Omega_1}{\partial x} + \cdots + \frac{\partial\Omega_1}{\partial y^{(n-2)}}y^{(n-1)} + \phi_1 y^{(n)}.$$

Then
$$\Psi - \frac{d\Omega_1}{dx} = \phi_3\left[\frac{d^{n-k}y}{dx^{n-k}}\right]^m + \phi_4.$$

That is, the expression $\Psi - \Omega_1'$ does not contain $y^{(n)}$ and may contain no derivative of order higher than $n - k$, and may be collected as

indicated. Now if Ψ was an exact derivative, so must $\Psi - \Omega_1'$ be. Hence if $m \neq 1$, the conclusion is that Ψ was not exact. If $m = 1$, the process of integration may be continued to obtain Ω_2 by integrating partially with respect to $y^{(n-k-1)}$. And so on until it is shown that Ψ is not exact or until Ψ is seen to be the derivative of an expression $\Omega_1 + \Omega_2 + \cdots = C$.

As an example consider $\Psi = x^2y''' + xy'' + (2xy - 1)y' + y^2 = 0$. Then

$$\Omega_1 = \int x^2 dy'' = x^2 y'', \qquad \Psi - \Omega_1' = -xy'' + (2xy - 1)y' + y^2,$$

$$\Omega_2 = \int -x dy' = -xy', \qquad \Psi - \Omega_1' - \Omega_2' = 2xyy' + y^2 = (xy^2)'.$$

As the expression of the first order is an exact derivative, the result is

$$\Psi - \Omega_1' - \Omega_2' - (xy^2)' = 0; \quad \text{and} \quad \Psi_1 = x^2y'' - xy' + xy^2 - C_1 = 0$$

is the new equation. The method may be tried again.

$$\Omega_1 = \int x^2 dy' = x^2 y', \qquad \Psi_1 - \Omega_1' = -3xy' + xy^2 - C_1.$$

This is not an exact derivative and the equation $\Psi_1 = 0$ is not exact. Moreover the equation $\Psi_1 = 0$ contains both x and y and is not homogeneous of any type except when $C_1 = 0$. It therefore appears as though the further integration of the equation $\Psi = 0$ were impossible.

The method is applied with especial ease to the case of

$$X_0 \frac{d^n y}{dx^n} + X_1 \frac{d^{n-1} y}{dx^{n-1}} + \cdots + X_{n-1} \frac{dy}{dx} + X_n y - R(x) = 0, \tag{17}$$

where the coefficients are functions of x alone. This is known as the *linear equation*, the integration of which has been treated only when the order is 1 or when the coefficients are constants. The application of successive integration by parts gives

$$\Omega_1 = X_0 y^{(n-1)}, \quad \Omega_2 = (X_1 - X_0')y^{(n-2)}, \quad \Omega_3 = (X_2 - X_1' + X_0'')y^{(n-3)}, \cdots;$$

and after n such integrations there is left merely

$$(X_n - X_{n-1}' + \cdots + (-1)^{n-1} X_1 + (-1)^n X_0)y - R,$$

which is a derivative only when it is a function of x. Hence

$$X_n - X_{n-1}' + \cdots + (-1)^{n-1} X_1 + (-1)^n X_0 = 0 \tag{18}$$

is the condition that the linear equation shall be exact, and

$$X_0 y^{(n-1)} + (X_1 - X_0')y^{(n-2)} + (X_2 - X_1' + X_0'')y^{(n-3)} + \cdots = \int R dx \tag{19}$$

is the first solution in case it is exact.

As an example take $y''' + y'' \cos x - 2y' \sin x - y \cos x = \sin 2x$. The test

$$X_3 - X_2' + X_1'' - X_0''' = -\cos x + 2\cos x - \cos x = 0$$

is satisfied. The integral is therefore $y'' + y'\cos x - y\sin x = -\frac{1}{2}\cos 2x + C_1$. This equation still satisfies the test for exactness. Hence it may be integrated again with the result $y' + y\cos x = -\frac{1}{4}\sin 2x + C_1x + C_2$. This belongs to the linear type. The final result is therefore

$$y = e^{-\sin x}\int e^{\sin x}(C_1x + C_2)\,dx + C_3e^{-\sin x} + \tfrac{1}{2}(1 - \sin x).$$

EXERCISES

1. Integrate these equations or at least reduce them to quadratures:

(α) $2xy'''y'' = y''^2 - a^2$, (β) $(1 + x^2)y'' + 1 + y'^2 = 0$,
(γ) $y^{\rm iv} + a^2y'' = 0$, (δ) $y^{\rm v} - m^2y''' = e^{ax}$, (ϵ) $x^2y^{\rm iv} + a^2y'' = 0$,
(ζ) $a^2y''y' = x$, (η) $xy'' + y' = 0$, (θ) $y'''y'' = 4$,
(ι) $(1 - x^2)y'' - xy' = 2$, (κ) $y^{\rm iv} = \sqrt{y'''}$, (λ) $y'' = f(y)$,
(μ) $2(2a - y)y'' = 1 + y'^2$, (ν) $yy'' - y'^2 - y^2y' = 0$,
(o) $yy'' + y'^2 + 1 = 0$, (π) $2y'' = e^y$, (ρ) $y^3y'' = a$.

2. Carry the integration as far as possible in these cases:

(α) $x^2y'' = (mx^2y'^2 + ny^2)^{\frac{1}{2}}$, (β) $mx^3y'' = (y - xy')^2$,
(γ) $x^4y'' = (y - xy')^3$, (δ) $x^4y'' - x^3y' - x^2y'^2 + 4y^2 = 0$,
(ϵ) $x^{-2}y'' + x^{-4}y = \frac{1}{4}y'^2$, (ζ) $ayy'' + by'^2 = yy'(c^2 + x^2)^{-\frac{1}{2}}$.

3. Carry the integration as far as possible in these cases:

(α) $(y^2 + x)y''' + 6yy'y'' + y'' + 2y'^2 = 0$, (β) $y'y'' - yx^2y' = xy^2$,
(γ) $x^3yy''' + 3x^3y'y'' + 9x^2yy'' + 9x^2y'^2 + 18xyy' + 3y^2 = 0$,
(δ) $y + 3xy' + 2yy'^3 + (x^2 + 2y^2y')y'' = 0$,
(ϵ) $(2x^3y' + x^2y)y'' + 4x^2y'^2 + 2xyy' = 0$.

4. Treat these linear equations:

(α) $xy'' + 2y = 2x$, (β) $(x^2 - 1)y'' + 4xy' + 2y = 2x$,
(γ) $y'' - y'\cot x + y\csc^2 x = \cos x$, (δ) $(x^2 - x)y'' + (3x - 2)y' + y = 0$,
(ϵ) $(x - x^3)y''' + (1 - 5x^2)y'' - 2xy' + 2y = 6x$,
(ζ) $(x^3 + x^2 - 3x + 1)y''' + (9x^2 + 6x - 9)y'' + (18x + 6)y' + 6y = x^3$,
(η) $(x + 2)^2y''' + (x + 2)y'' + y' = 1$, (θ) $x^2y'' + 3xy' + y = x$,
(ι) $(x^3 - x)y''' + (8x^2 - 3)y'' + 14xy' + 4y = 0$.

5. Note that Ex. 4 (θ) comes under the third homogeneous type, and that Ex. 4 (η) may be brought under that type by multiplying by $(x + 2)$. Test sundry of Exs. 1, 2, 3 for exactness. Show that any linear equation in which the coefficients are polynomials of degree less than the order of the derivatives of which they are the coefficients, is surely exact.

6. Sometimes, when the condition that an equation be exact is not satisfied, it is possible to find an integrating factor for the equation so that after multiplication by the factor the equation becomes exact. For linear equations try x^m. Integrate

(α) $x^5y'' + (2x^4 - x)y' - (2x^3 - 1)y = 0$. (β) $(x^2 - x^4)y'' - x^3y' - 2y = 0$.

7. Show that the equation $y'' + Py' + Qy'^2 = 0$ may be reduced to quadratures 1° when P and Q are both functions of y, or 2° when both are functions of x, or 3° when P is a function of x and Q is a function of y (integrating factor $1/y'$). In each case find the general expression for y in terms of quadratures. Integrate $y'' + 2y'\cot x + 2y'^2\tan y = 0$.

8. Find and discuss the curves for which the radius of curvature is proportional to the radius r of the curve.

9. If the radius of curvature R is expressed as a function $R = R(s)$ of the arc s measured from some point, the equation $R = R(s)$ or $s = s(R)$ is called the *intrinsic equation* of the curve. To find the relation between x and y the second equation may be differentiated as $ds = s'(R)\,dR$, and this equation of the third order may be solved. Show that if the origin be taken on the curve at the point $s = 0$ and if the x-axis be tangent to the curve, the equations

$$x = \int_0^s \cos\left[\int_0^s \frac{ds}{R}\right] ds, \qquad y = \int_0^s \sin\left[\int_0^s \frac{ds}{R}\right] ds$$

express the curve parametrically. Find the curves whose intrinsic equations are

$$(\alpha)\ R = a, \qquad (\beta)\ aR = s^2 + a^2, \qquad (\gamma)\ R^2 + s^2 = 16a^2.$$

10. Given $F = y^{(n)} + X_1 y^{(n-1)} + X_2 y^{(n-2)} + \cdots + X_{n-1} y' + X_n y = 0$. Show that if μ, a function of x alone, is an integrating factor of the equation, then

$$\Phi = \mu^{(n)} - (X_1\mu)^{(n-1)} + (X_2\mu)^{(n-2)} - \cdots + (-1)^{n-1}(X_{n-1}\mu)' + (-1)^n X_n\mu = 0$$

is the equation satisfied by μ. Collect the coefficient of μ to show that the condition that the given equation be exact is the condition that this coefficient vanish. The equation $\Phi = 0$ is called the *adjoint* of the given equation $F = 0$. Any integral μ of the adjoint equation is an integrating factor of the original equation. Moreover note that

$$\int \mu F dx = \mu y^{(n-1)} + (\mu X_1 - \mu')\,y^{(n-2)} + \cdots + (-1)^n \int y\Phi dx,$$

or $$d[\mu F - (-1)^n y\Phi] = d\,[\mu y^{(n-1)} + (\mu X_1 - \mu')\,y^{(n-2)} + \cdots] = d\Omega.$$

Hence if μF is an exact differential, so is $y\Phi$. In other words, any solution y of the original equation is an integrating factor for the adjoint equation.

104. Linear differential equations. The equations

$$\begin{aligned} X_0 D^n y + X_1 D^{n-1} y + \cdots + X_{n-1} Dy + X_n y &= R(x), \\ X_0 D^n y + X_1 D^{n-1} y + \cdots + X_{n-1} Dy + X_n y &= 0 \end{aligned} \tag{20}$$

are linear differential equations of the nth order; the first is called the *complete equation* and the second the *reduced equation*. If $y_1, y_2, y_3, \cdots$ are any solutions of the reduced equation, and $C_1, C_2, C_3, \cdots$ are any constants, then $y = C_1 y_1 + C_2 y_2 + C_3 y_3 + \cdots$ is also a solution of the reduced equation. This follows at once from the linearity of the reduced equation and is proved by direct substitution. Furthermore if I is any solution of the complete equation, then $y + I$ is also a solution of the complete equation (cf. § 96).

As the equations (20) are of the nth order, they will determine $y^{(n)}$ and, by differentiation, all higher derivatives in terms of the values of $x, y, y', \cdots, y^{(n-1)}$. Hence if the values of the n quantities $y_0, y_0', \cdots, y_0^{(n-1)}$ which correspond to the value $x = x_0$ be given, all the higher derivatives are determined (§§ 87–88). Hence there are n and no more than n arbitrary conditions that may be imposed as initial conditions. A solution

of the equations (20) which contains n distinct arbitrary constants is called the general solution. By distinct is meant that the constants can actually be determined to suit the n initial conditions.

If $y_1, y_2, \cdots, y_n$ are n solutions of the reduced equation, and

$$\begin{aligned} y &= C_1y_1 + C_2y_2 + \cdots + C_ny_n, \\ y' &= C_1y_1' + C_2y_2' + \cdots + C_ny_n', \\ &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ y^{(n-1)} &= C_1y_1^{(n-1)} + C_2y_2^{(n-1)} + \cdots + C_ny_n^{(n-1)}, \end{aligned} \tag{21}$$

then y is a solution and $y', \cdots, y^{(n-1)}$ are its first $n-1$ derivatives. If x_0 be substituted on the right and the assumed corresponding initial values $y_0, y_0', \cdots, y_0^{(n-1)}$ be substituted on the left, the above n equations become linear equations in the n unknowns $C_1, C_2, \cdots, C_n$; and if they are to be soluble for the C's, the condition

$$W(y_1, y_2, \cdots, y_n) = \begin{vmatrix} y_1 & y_2 & \cdots & y_n \\ y_1' & y_2' & \cdots & y_n' \\ \cdot & \cdot & \cdot & \cdot \\ y_1^{(n-1)} & y_2^{(n-1)} & \cdots & y_n^{(n-1)} \end{vmatrix} \neq 0 \tag{22}$$

must hold for every value of $x = x_0$. Conversely if the condition does hold, the equations will be soluble for the C's.

The determinant $W(y_1, y_2, \cdots, y_n)$ is called the *Wronskian* of the n functions $y_1, y_2, \cdots, y_n$. The result may be stated as: If n functions $y_1, y_2, \cdots, y_n$ which are solutions of the reduced equation, and of which the Wronskian does not vanish, can be found, the general solution of the reduced equation can be written down. In general no solution of the equation can be found, whether by a definite process or by inspection; but in the rare instances in which the n solutions can be seen by inspection the problem of the solution of the reduced equation is completed. Frequently one solution may be found by inspection, and it is therefore important to see how much this contributes toward effecting the solution.

If y_1 is a solution of the reduced equation, make the substitution $y = y_1z$. The derivatives of y may be obtained by Leibniz's Theorem (§ 8). As the formula is linear in the derivatives of z, it follows that the result of the substitution will leave the equation linear in the new variable z. Moreover, to collect the coefficient of z itself, it is necessary to take only the first term $y_1^{(k)}z$ in the expansions for the derivative $y^{(k)}$. Hence

$$(X_0y_1^{(n)} + X_1y_1^{(n-1)} + \cdots + X_{n-1}y_1' + X_ny_1)z = 0$$

is the coefficient of z and vanishes by the assumption that y_1 is a solution of the reduced equation. Then the equation for z is

$$P_0z^{(n)} + P_1z^{(n-1)} + \cdots + P_{n-2}z'' + P_{n-1}z' = 0; \tag{23}$$

and if z' be taken as the variable, the equation is of the order $n-1$. It therefore appears that *the knowledge of a solution y_1 reduces the order of the equation by one.*

Now if $y_2, y_3, \cdots, y_p$ were other solutions, the derived ratios

$$z_1' = \left(\frac{y_2}{y_1}\right)', \qquad z_2' = \left(\frac{y_3}{y_1}\right)', \qquad \cdots, \qquad z_{p-1}' = \left(\frac{y_p}{y_1}\right)' \tag{23'}$$

would be solutions of the equation in z'; for by substitution,

$$y = y_1 z_1 = y_2, \qquad y = y_1 z_2 = y_3, \qquad \cdots, \qquad y = y_1 z_{p-1} = y_p$$

are all solutions of the equation in y. Moreover, if there were a linear relation $C_1 z_1' + C_2 z_2' + \cdots + C_{p-1} z_{p-1}' = 0$ connecting the solutions z_i', an integration would give a linear relation

$$C_1 y_2 + C_2 y_3 + \cdots + C_{p-1} y_n + C_p y_1 = 0$$

connecting the p solutions y_i. Hence if there is no linear relation (of which the coefficients are not all zero) connecting the p solutions y_i of the original equation, there can be none connecting the $p-1$ solutions z_i' of the transformed equation. Hence *a knowledge of p solutions of the original reduced equation gives a new reduced equation of which $p-1$ solutions are known.* And the process of substitution may be continued to reduce the order further until the order $n-p$ is reached.

As an example consider the equation of the third order

$$(1-x)y''' + (x^2-1)y'' - x^2y' + xy = 0.$$

Here a simple trial shows that x and e^x are two solutions. Substitute

$$y = e^x z, \qquad y' = e^x(z+z'), \qquad y'' = e^x(z+2z'+z''), \qquad y''' = e^x(z+3z'+3z''+z''').$$

Then $$(1-x)z''' + (x^2-3x+2)z'' + (x^2-3x+1)z' = 0$$

is of the second order in z'. A known solution is the derived ratio $(x/e^x)'$.

$$z' = (xe^{-x})' = e^{-x}(1-x). \quad \text{Let } z' = e^{-x}(1-x)w.$$

From this, z'' and z''' may be found and the equation takes the form

$$(1-x)w'' + (1+x)(x-2)w' = 0 \quad \text{or} \quad \frac{dw'}{w'} = xdx - \frac{2}{x-1}dx.$$

This is a linear equation of the first order and may be solved.

$$\log w' = \tfrac{1}{2}x^2 - 2\log(x-1) + C \quad \text{or} \quad w' = C_1 e^{\frac{1}{2}x^2}(x-1)^{-2}.$$

Hence $$w = C_1 \int e^{\frac{1}{2}x^2}(x-1)^{-2}dx + C_2,$$

$$z' = \left(\frac{x}{e^x}\right)' w = C_1 \left(\frac{x}{e^x}\right)' \int e^{\frac{1}{2}x^2}(x-1)^{-2}dx + C_2\left(\frac{x}{e^x}\right)',$$

$$z = C_1 \int \left(\frac{x}{e^x}\right)' \int e^{\frac{1}{2}x^2}(x-1)^{-2}(dx)^2 + C_2\frac{x}{e^x} + C_3,$$

$$y = e^x z = C_1 e^x \int \left(\frac{x}{e^x}\right)' \int e^{\frac{1}{2}x^2}(x-1)^{-2}(dx)^2 + C_2 x + C_3 e^x.$$

The value for y is thus obtained in terms of quadratures. It may be shown that in case the equation is of the nth degree with p known solutions, the final result will call for $p(n-p)$ quadratures.

105. If the general solution $y = C_1y_1 + C_2y_2 + \cdots + C_ny_n$ of the reduced equation has been found (called the *complementary function* for the complete equation), the general solution of the complete equation may always be obtained in terms of quadratures by the important and far-reaching *method of the variation of constants* due to Lagrange. The question is: Cannot functions of x be found so that the expression

$$y = C_1(x)y_1 + C_2(x)y_2 + \cdots + C_n(x)y_n \tag{24}$$

shall be the solution of the complete equation? As there are n of these functions to be determined, it should be possible to impose $n-1$ conditions upon them and still find the functions.

Differentiate y on the supposition that the C's are variable.

$$y' = C_1y_1' + C_2y_2' + \cdots + C_ny_n' + y_1C_1' + y_2C_2' + \cdots + y_nC_n'.$$

As one of the conditions on the C's suppose that

$$y_1C_1' + y_2C_2' + \cdots + y_nC_n' = 0.$$

Differentiate again and impose the new condition

$$y_1'C_1' + y_2'C_2' + \cdots + y_n'C_n' = 0,$$

so that $$y'' = C_1y_1'' + C_2y_2'' + \cdots + C_ny_n''.$$

The differentiation may be continued to the $(n-1)$st condition

$$y_1^{(n-2)}C_1' + y_2^{(n-2)}C_2' + \cdots + y_n^{(n-2)}C_n' = 0,$$

and $$y^{(n-1)} = C_1y_1^{(n-1)} + C_2y_2^{(n-1)} + \cdots + C_ny_n^{(n-1)}.$$

Then $$y^{(n)} = C_1y_1^{(n)} + C_2y_2^{(n)} + \cdots + C_ny_n^{(n)} + y_1^{(n-1)}C_1' + y_2^{(n-1)}C_2' + \cdots + y_n^{(n-1)}C_n'.$$

Now if the expressions thus found for $y, y', y'', \cdots, y^{(n-1)}, y^{(n)}$ be substituted in the complete equation, and it be remembered that $y_1, y_2, \cdots, y_n$ are solutions of the reduced equation and hence give 0 when substituted in the left-hand side of the equation, the result is

$$y_1^{(n-1)}C_1' + y_2^{(n-1)}C_2' + \cdots + y_n^{(n-1)}C_n' = R.$$

Hence, in all, there are n linear equations

$$\begin{aligned}
y_1C_1' \quad &+ y_2C_2' \quad + \cdots + y_nC_n' \quad = 0,\\
y_1'C_1' \quad &+ y_2'C_2' \quad + \cdots + y_n'C_n' \quad = 0,\\
&\cdots\cdots\cdots\cdots\\
y_1^{(n-2)}C_1' &+ y_2^{(n-2)}C_2' + \cdots + y_n^{(n-2)}C_n' = 0,\\
y_1^{(n-1)}C_1' &+ y_2^{(n-1)}C_2' + \cdots + y_n^{(n-1)}C_n' = R.
\end{aligned} \tag{25}$$

connecting the derivatives of the C's; and these may actually be solved for those derivatives which will then be expressed in terms of x. The C's may then be found by quadrature.

As an example consider the equation with constant coefficients

$$(D^3 + D)y = \sec x \quad \text{with} \quad y = C_1 + C_2 \cos x + C_3 \sin x$$

as the solution of the reduced equation. Here the solutions y_1, y_2, y_3 may be taken as 1, $\cos x$, $\sin x$ respectively. The conditions on the derivatives of the C's become by direct substitution in (25)

$$C_1' + \cos x C_2' + \sin x C_3' = 0, \quad -\sin x C_2' + \cos x C_3' = 0, \quad -\cos x C_2' - \sin x C_3' = \sec x.$$

Hence $\qquad C_1' = \sec x, \qquad C_2' = -1, \qquad C_3' = -\tan x$

and $\quad C_1 = \log\tan(\tfrac{1}{2}x + \tfrac{1}{4}\pi) + c_1, \qquad C_2 = -x + c_2, \qquad C_3 = \log\cos x + c_3.$

Hence $\qquad y = c_1 + \log\tan(\tfrac{1}{2}x + \tfrac{1}{4}\pi) + (c_2 - x)\cos x + (c_3 + \log\cos x)\sin x$

is the general solution of the complete equation. This result could not be obtained by any of the real short methods of §§ 96–97. It could be obtained by the general method of § 95, but with little if any advantage over the method of variation of constants here given. The present method is equally available for equations with variable coefficients.

106. *Linear equations of the second order* are especially frequent in practical problems. In a number of cases the solution may be found. Thus 1° when the coefficients are constant or may be made constant by a change of variable as in Ex. 7, p. 222, the general solution of the reduced equation may be written down at once. The solution of the complete equation may then be found by obtaining a particular integral I by the methods of §§ 95–97 or by the application of the method of variation of constants. And 2° when the equation is exact, the solution may be had by integrating the linear equation (19) of § 103 of the first order by the ordinary methods. And 3° when one solution of the reduced equation is known (§ 104), the reduced equation may be completely solved and the complete equation may then be solved by the method of variation of constants, or the complete equation may be solved directly by Ex. 6 below.

Otherwise, write the differential equation in the form

$$\frac{d^2y}{dx^2} + P\frac{dy}{dx} + Qy = R. \tag{26}$$

The substitution $y = uz$ gives the new equation

$$\frac{d^2z}{dx^2} + \left(\frac{2}{u}\frac{du}{dx} + P\right)\frac{dz}{dx} + \frac{1}{u}(u'' + Pu' + Qu)z = \frac{R}{u}. \tag{26'}$$

If u be determined so that the coefficient of z' vanishes, then

$$u = e^{-\frac{1}{2}\int P dx} \quad \text{and} \quad \frac{d^2z}{dx^2} + \left(Q - \frac{1}{2}\frac{dP}{dx} - \frac{1}{4}P^2\right)z = Re^{\frac{1}{2}\int P dx}. \tag{27}$$

Now 4° if $Q - \frac{1}{2}P' - \frac{1}{4}P^2$ is constant, the new reduced equation in (27) may be integrated; and 5° if it is k/x^2, the equation may also be integrated by the method of Ex. 7, p. 222. The integral of the complete equation may then be found. (In other cases this method may be useful in that the equation is reduced to a simpler form where solutions of the reduced equation are more evident.)

Again, suppose that the independent variable is changed to z. Then

$$\frac{d^2y}{dz^2} + \frac{z'' + Pz'}{z'^2}\frac{dy}{dz} + \frac{Q}{z'^2}y = \frac{R}{z'^2}. \tag{28}$$

Now 6° if $z'^2 = \pm Q$ will make $z'' + Pz' = kz'^2$, so that the coefficient of dy/dz becomes a constant k, the equation is integrable. (Trying if $z'^2 = \pm Qz^2$ will make $z'' + Pz' = kz'^2/z$ is needless because nothing in addition to 6° is thereby obtained. It may happen that if z be determined so as to make $z'' + Pz' = 0$, the equation will be so far simplified that a solution of the reduced equation becomes evident.)

Consider the example $\frac{d^2y}{dx^2} + \frac{2}{x}\frac{dy}{dx} + \frac{a^2}{x^4}y = 0$. Here no solution is apparent. Hence compute $Q - \frac{1}{2}P' - \frac{1}{4}P^2$. This is a^2/x^4 and is neither constant nor proportional to $1/x^2$. Hence the methods 4° and 5° will not work. From $z'^2 = Q = a^2/x^4$ or $z' = a/x^2$, it appears that $z'' + Pz' = 0$, and 6° works; the new equation is

$$\frac{d^2y}{dz^2} + y = 0 \quad \text{with} \quad z = -\frac{a}{x}.$$

The solution is therefore seen immediately to be

$$y = C_1 \cos z - C_2 \sin z \quad \text{or} \quad y = C_1 \cos(a/x) + C_2 \sin(a/x).$$

If there had been a right-hand member in the original equation, the solution could have been found by the method of variation of constants, or by some of the short methods for finding a particular solution if R had been of the proper form.

EXERCISES

1. If a relation $C_1y_1 + C_2y_2 + \cdots + C_ny_n = 0$, with constant coefficients not all 0, exists between n functions $y_1, y_2, \cdots, y_n$ of x for all values of x, the functions are by definition said to be *linearly dependent;* if no such relation exists, they are said to be *linearly independent.* Show that the nonvanishing of the Wronskian is a criterion for linear independence.

2. If the general solution $y = C_1y_1 + C_2y_2 + \cdots + C_ny_n$ is the same for

$$X_0y^{(n)} + X_1y^{(n-1)} + \cdots + X_ny = 0 \quad \text{and} \quad P_0y^{(n)} + P_1y^{(n-1)} + \cdots + P_ny = 0,$$

two linear equations of the nth order, show that y satisfies the equation

$$(X_1P_0 - X_0P_1)y^{(n-1)} + \cdots + (X_nP_0 - X_0P_n)y = 0$$

of the $(n-1)$st order; and hence infer, from the fact that y contains n arbitrary constants corresponding to n arbitrary initial conditions, the important theorem: If two linear equations of the nth order have the same general solution, the corresponding coefficients are proportional.

3. If $y_1, y_2, \cdots, y_n$ are n independent solutions of an equation of the nth order, show that the equation may be taken in the form $W(y_1, y_2, \cdots, y_n, y) = 0$.

4. Show that if, in any reduced equation, $X_{n-1} + xX_n = 0$ identically, then x is a solution. Find the condition that x^m be a solution; also that e^{mx} be a solution.

5. Find by inspection one or more independent solutions and integrate:

(α) $(1+x^2)y'' - 2xy' + 2y = 0$, ($\beta$) $xy'' + (1-x)y' - y = 0$,

(γ) $(ax - bx^2)y'' - ay' + 2by = 0$, ($\delta$) $\frac{1}{2}y'' + xy' - (x+2)y = 0$,

(ϵ) $\left(\log x + \frac{1}{x^4} - \frac{1}{x^2} + \frac{1}{x}\right)y''' + \left(\log x + \frac{1}{x^4} + \frac{1}{x^3} - \frac{1}{x^2}\right)y'' + \left(\frac{1}{x^2} - \frac{1}{x}\right)(y' - xy) = 0$.

(ζ) $y^{\text{iv}} - xy''' + xy' - y = 0$. ($\eta$) $(4x^2 - x + 1)y''' + 8x^2y'' - 4xy' - 8y = 0$.

6. If y_1 is a known solution of the equation $y'' + Py' + Qy = R$ of the second order, show that the general solution may be written as

$$y = C_1y_1 + C_2y_1\int e^{-\int Pdx}\frac{dx}{y_1^2} + y_1\int\frac{1}{y_1^2}e^{-\int Pdx}\int y_1e^{\int Pdx}R\,(dx)^2.$$

7. Integrate:

(α) $xy'' - (2x+1)y' + (x+1)y = x^2 - x - 1$.

(β) $y'' - x^2y' + xy = x$, (γ) $xy'' + (1-x)y' - y = e^x$,

(δ) $y'' - xy' + (x-1)y = R$, (ϵ) $y''\sin^2 x + y'\sin x\cos x - y = x - \sin x$.

8. After writing down the integral of the reduced equation by inspection, apply the method of the variation of constants to these equations:

(α) $(D^2+1)y = \tan x$, (β) $(D^2+1)y = \sec^2 x$. (γ) $(D-1)^2y = e^x(1-x)^{-2}$,

(δ) $(1-x)y'' + xy' - y = (1-x)^2$, ($\epsilon$) $(1 - 2x + x^2)(y''' - 1) - x^2y'' + 2xy' - y = 1$.

9. Integrate the following equations of the second order:

(α) $4x^2y'' + 4x^3y' + (x^2+1)^2y = 0$. ($\beta$) $y'' - 2y'\tan x - (a^2+1)y = 0$.

(γ) $xy'' + 2y' - xy = 2e^x$, (δ) $y''\sin x + 2y'\cos x + 3y\sin x = e^x$.

(ϵ) $y'' + y'\tan x + y\cos^2 x = 0$, ($\zeta$) $(1-x^2)y'' - xy' + 4y = 0$.

(η) $y'' + (2e^x - 1)y' + e^{2x}y = e^{4x}$. ($\theta$) $x^6y'' + 3x^5y' + y = x^{-2}$.

10. Show that if $X_0y'' + X_1y' + X_2y = R$ may be written in factors as

$$(X_0D^2 + X_1D + X_2)y = (p_1D + q_1)(p_2D + q_2)y = R,$$

where the factors are not commutative inasmuch as the differentiation in one factor is applied to the variable coefficients of the succeeding factor as well as to D, then the solution is obtainable in terms of quadratures. Show that

$$q_1p_2 + p_1p_2' + p_1q_2 = X_1 \quad \text{and} \quad q_1q_2 + p_1q_2' = X_2.$$

In this manner integrate the following equations, choosing p_1 and p_2 as factors of X_0 and determining q_1 and q_2 by inspection or by assuming them in some form and applying the method of undetermined coefficients:

(α) $xy'' + (1-x)y' - y = e^x$. (β) $3x^2y'' + (2 - 6x^2)y' - 4 = 0$.

(γ) $3x^2y'' + (2 + 6x - 6x^2)y' - 4y = 0$. ($\delta$) $(x^2-1)y'' - (3x+1)y' - x(x-1)y = 0$.

(ϵ) $axy'' + (3a + bx)y' + 3by = 0$. ($\zeta$) $xy'' - 2x(1+x)y' + 2(1+x)y = x^3$.

11. Integrate these equations in any manner:

(α) $y'' - \frac{1}{\sqrt{x}}y' + \frac{x + \sqrt{x} - 8}{4x^2}y = 0$. ($\beta$) $y'' - \frac{2}{x}y' + \left(a^2 + \frac{2}{x^2}\right)y = 0$,

(γ) $y'' + y' \tan x + y \cos^2 x = 0$, ($\delta$) $y'' - 2\left(n - \frac{a}{x}\right) y' + \left(n^2 - 2\frac{na}{x}\right) y = e^{nx}$,

(ϵ) $(1 - x^2) y'' - xy' - c^2 y = 0$, ($\zeta$) $(a^2 - x^2) y'' - 8xy' - 12y = 0$,

(η) $y'' + \frac{1}{x^2 \log x} y = e^x\left(\frac{2}{x} + \log x\right)$, ($\theta$) $y'' - \frac{9 - 4x}{3 - x} y' + \frac{6 - 3x}{3 - x} y = 0$,

(ι) $y'' + 2x^{-1} y' - n^2 y = 0$, ($\kappa$) $y'' - 4xy' + (4x^2 - 3) y = e^{x^2}$,

(λ) $y'' + 2ny' \cot nx + (m^2 - n^2) y = 0$, ($\mu$) $y'' + 2(x^{-1} + Bx^{-2}) y' + Ax^{-4} y = 0$.

12. If y_1 and y_2 are solutions of $y'' + Py' + R = 0$, show by eliminating Q and integrating that

$$y_1 y_2' - y_2 y_1' = Ce^{-\int P dx}.$$

What if $C = 0$? If $C \neq 0$, note that y_1 and y_1' cannot vanish together; and if $y_1(a) = y_1(b) = 0$, use the relation $(y_2 y_1')_a : (y_2 y_1')_b = k > 0$ to show that as y_{1a}' and y_{1b}' have opposite signs, y_{2a} and y_{2b} have opposite signs and hence $y_2(\xi) = 0$ where $a < \xi < b$. Hence the theorem: Between any two roots of a solution of an equation of the second order there is one root of every solution independent of the given solution. What conditions of continuity for y and y' are tacitly assumed here?

107. The cylinder functions. Suppose that $C_n(x)$ is a function of x which is different for different values of n and which satisfies the two equations

$$C_{n-1}(x) - C_{n+1}(x) = 2\frac{d}{dx} C_n(x), \quad C_{n-1}(x) + C_{n+1}(x) = \frac{2n}{x} C_n(x). \tag{29}$$

Such a function is called a *cylinder function* and the index n is called the *order* of the function and may have any real value. The two equations are supposed to hold for all values of n and for all values of x. They do not completely determine the functions but from them follow the chief rules of operation with the functions. For instance, by addition and subtraction,

$$C_n'(x) = C_{n-1}(x) - \frac{n}{x} C_n(x) = \frac{n}{x} C_n(x) - C_{n+1}(x). \tag{30}$$

Other relations which are easily deduced are

$$D_x[x^n C_n(\alpha x)] = \alpha x^n C_{n-1}(\alpha x), \qquad D_x[x^{-n} C_n(\alpha x)] = -\alpha x^{-n} C_{n+1}(x), \tag{31}$$

$$D_x\left[x^{\frac{n}{2}} C_n(\sqrt{\alpha x})\right] = \tfrac{1}{2}\sqrt{\alpha}\, x^{\frac{n-1}{2}} C_{n-1}(\sqrt{\alpha x}), \tag{32}$$

$$C_0'(x) = -C_1(x), \qquad C_{-n}(x) = (-1)^n C_n(x), \qquad n \text{ integral}, \tag{33}$$

$$C_n(x) K_n'(x) - C_n'(x) K_n(x) = C_{n+1}(x) K_n(x) - C_n(x) K_{n+1}(x) = \frac{A}{x}, \tag{34}$$

where C and K denote any two cylinder functions.

The proof of these relations is simple, but will be given to show the use of (29). In the first case differentiate directly and substitute from (29).

$$D_x[x^n C_n(\alpha x)] = x^n\left[\alpha D_{\alpha x} C_n(\alpha x) + \frac{n}{x} C_n(\alpha x)\right]$$
$$= x^n\left[\alpha C_{n-1}(\alpha x) - \alpha\frac{n}{\alpha x} C_n(\alpha x) + \frac{n}{x} C_n(\alpha x)\right].$$

The second of (31) is proved similarly. For (32), differentiate.

$$D_x\left[x^{\frac{n}{2}}C_n(\sqrt{\alpha x})\right] = \frac{1}{2}nx^{\frac{n}{2}-1}C_n(\sqrt{\alpha x}) + x^{\frac{n}{2}}\frac{1}{2}\sqrt{\frac{\alpha}{x}}D_{\sqrt{\alpha x}}C_n(\sqrt{\alpha x})$$

$$= \frac{1}{2}\sqrt{\alpha}\,x^{\frac{n-1}{2}}\left[\frac{n}{\sqrt{\alpha x}}C_n(\sqrt{\alpha x}) + C_{n-1}(\sqrt{\alpha x}) - \frac{n}{\sqrt{\alpha x}}C_n(\sqrt{\alpha x})\right].$$

Next (33) is obtained 1° by substituting 0 for n in both equations (29).

$$C_{-1}(x) - C_1(x) = 2\,C_0'(x), \quad C_{-1}(x) + C_1(x) = 0, \quad \text{hence} \quad C_0'(x) = -C_1(x);$$

and 2° by substituting successive values for n in the second of (29) written in the form $xC_{n+1} + xC_{n-1} = 2\,nC_n$. Then

$$xC_{-1} + xC_1 = 0, \qquad xC_{-2} + xC_0 = -2\,C_{-1}, \qquad xC_0 + xC_2 = 2\,C_1,$$
$$xC_{-3} + xC_{-1} = -4\,C_{-2}, \qquad xC_1 + xC_3 = 4\,C_2,$$
$$xC_{-4} + xC_{-2} = -6\,C_{-3}, \qquad xC_2 + xC_4 = 6\,C_3,$$

and so on. The first gives $C_{-1} = -C_1$. Subtract the next two and use $C_{-1} + C_1 = 0$. Then $C_{-2} - C_2 = 0$ or $C_{-2} = (-1)^2C_2$. Add the next two and use the relations already found. Then $C_{-3} + C_3 = 0$ or $C_{-3} = (-1)^3C_3$. Subtract the next two, and so on. For the last of the relations, a very important one, note first that the two expressions become equivalent by virtue of (29); for

$$C_nK_n' - C_n'K_n = \frac{n}{x}C_nK_n - C_nK_{n+1} - \frac{n}{x}C_nK_n + C_{n+1}K_n.$$

Now $$\frac{d}{dx}\left[x(C_{n+1}K_n - C_nK_{n+1})\right] = C_{n+1}K_n - C_nK_{n+1} + xK_n\left(C_n - \frac{n+1}{x}C_{n+1}\right)$$
$$+ xC_{n+1}\left(\frac{n}{x}K_n - K_{n+1}\right) - xK_{n+1}\left(\frac{n}{x}C_n - C_{n+1}\right)$$
$$- xC_n\left(K_n - \frac{n+1}{x}K_{n+1}\right) = 0.$$

Hence $x(C_{n+1}K_n - C_nK_{n+1}) = \text{const.} = A$, and the relation is proved.

The cylinder functions of a given order n satisfy a linear differential equation of the second order. This may be obtained by differentiating the first of (29) and combining with (30).

$$2\,C_n'' = C_{n-1}' - C_{n+1}' = \frac{n-1}{x}C_{n-1} - 2\,C_n + \frac{n+1}{x}C_{n+1}$$
$$= \frac{n}{x}(C_{n-1} + C_{n+1}) - \frac{1}{x}(C_{n-1} - C_{n+1}) - 2\,C_n.$$

Hence $$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(1 - \frac{n^2}{x^2}\right)y = 0, \qquad y = C_n(x). \tag{35}$$

This equation is known as *Bessel's equation*; the functions $C_n(x)$, which have been called cylinder functions, are often called *Bessel's functions*. From the equation it follows that any three functions of the same order n are connected by a linear relation and there are only two independent functions of any given order.

By a change of the independent variable, the Bessel equation may take on several other forms. The easiest way to find them is to operate directly with the relations (31), (32). Thus

$$D_x[x^{-n}C_n(x)] = -x^{-n}C_{n+1} = -x \cdot x^{-n-1}C_{n+1},$$
$$D_x^2[x^{-n}C_n(x)] = -x^{-n-1}C_{n+1} + x \cdot x^{-n-1}C_{n+2}$$
$$= -x^{-n-1}C_{n+1} + 2(n+1)x^{-n-1}C_{n+1} - x^{-n}C_n.$$

Hence $$\frac{d^2y}{dx^2} + \frac{(1+2n)}{x}\frac{dy}{dx} + y = 0, \qquad y = x^{-n}C_n(x). \tag{36}$$

Again $$\frac{d^2y}{dx^2} + \frac{(1-2n)}{x}\frac{dy}{dx} + y = 0, \qquad y = x^{n}C_n(x). \tag{37}$$

Also $$xy'' + (1+n)y' + y = 0, \qquad y = x^{-\frac{n}{2}}C_n(2\sqrt{x}). \tag{38}$$

And $$xy'' + (1-n)y' + y = 0, \qquad y = x^{\frac{n}{2}}C_n(2\sqrt{x}). \tag{39}$$

In all these differential equations it is well to restrict x to positive values inasmuch as, if n is not specialized, the powers of x, as x^n, x^{-n}, $x^{\frac{n}{2}}$, $x^{-\frac{n}{2}}$, are not always real.

108. The fact that n occurs only squared in (35) shows that both $C_n(x)$ and $C_{-n}(x)$ are solutions, so that if these functions are independent, the complete solution is $y = aC_n + bC_{-n}$. In like manner the equations (36), (37) form a pair which differ only in the sign of n. Hence if H_n and H_{-n} denote particular integrals of the first and second respectively, the complete integrals are respectively

$$y = aH_n + bH_{-n}x^{-2n} \quad \text{and} \quad y = aH_{-n} + bH_n x^{2n};$$

and similarly the respective integrals of (38), (39) are

$$y = aI_n + bI_{-n}x^{-n} \quad \text{and} \quad y = aI_{-n} + bI_n x^n,$$

where I_n and I_{-n} denote particular integrals of these two equations. It should be noted that these forms are the complete solutions only when the two integrals are independent. Note that

$$I_n(x) = x^{-\frac{1}{2}n}C_n(2\sqrt{x}), \qquad C_n(x) = (\tfrac{1}{2}x)^n I_n(\tfrac{1}{4}x^2). \tag{40}$$

As it has been seen that $C_n = (-1)^n C_{-n}$ when n is integral, it follows that in this case the above forms do not give the complete solution.

A particular solution of (38) may readily be obtained in series by the method of undetermined coefficients (§ 88). It is

$$I_n(x) = \sum_0^\infty a_i x^i, \qquad a_i = \frac{(-1)^i}{i!(n+1)(n+2)\cdots(n+i)}, \tag{41}$$

as is derived below. It should be noted that I_{-n} formed by changing the sign of n is meaningless when n is an integer, for the reason that

from a certain point on, the coefficients a_i have zeros in the denominator. The determination of a series for the second independent solution when n is integral will be omitted. The solutions of (35), (36) corresponding to $I_n(x)$ are, by (40) and (41),

$$J_n(x) = \frac{x^n}{2^n} \sum_0^\infty \frac{(-1)^i x^{2i}}{2^{2i} i!(n+i)!} = \frac{x^n}{2^n n!} I_n(\tfrac{1}{4} x^2), \tag{42}$$

$$x^{-n} J_n(x) = \frac{1}{2^n n!} I_n(\tfrac{1}{4} x^2), \tag{42'}$$

where the factor $n!$ has been introduced in the denominator merely to conform to usage.* The chief cylinder function $C_n(x)$ is $J_n(x)$ and it always carries the name of Bessel.

To derive the series for $I_n(x)$ write

$$\begin{array}{r|l} 1 & I_n = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1} + \cdots, \\ (1+n) & I_n' = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + (k-1) a_{k-1} x^{k-2} + \cdots, \\ x & I_n'' = 2a_2 + 3 \cdot 2 a_3 x + \cdots + (k-1)(k-2) a_{k-1} x^{k-3} + \cdots, \\ \hline & 0 = [a_0 + a_1(n+1)] + x[a_1 + a_2 2(n+2)] + x^2[a_2 + a_3 3(n+3)] \\ & \qquad + \cdots + x^{k-1}[a_{k-1} + a_k k(n+k)] + \cdots. \end{array}$$

Hence $a_0 + a_1(n+1) = 0, \quad a_1 + a_2 2(n+2) = 0, \cdots, \quad a_{k-1} + a_k k(n+k) = 0,$

or $$a_1 = -\frac{a_0}{n+1}, \qquad a_2 = \frac{-a_1}{2(n+2)} = \frac{a_0}{2!(n+1)(n+2)}, \cdots,$$

$$a_k = \frac{(-1)^k a_0}{k!(n+1)\cdots(n+k)}.$$

If now the choice $a_0 = 1$ is made, the series for $I_n(x)$ is as given in (41).

The famous differential equation of the first order

$$xy' - ay + by^2 = cx^n, \tag{43}$$

known as *Riccati's equation*, may be integrated in terms of cylinder functions. Note that if $n = 0$ or $c = 0$, the variables are separable; and if $b = 0$, the equation is linear. As these cases are immediately integrable, assume $bcn \neq 0$. By a suitable change of variable, the equation takes the form

$$\xi \frac{d^2\eta}{d\xi^2} + \left(1 - \frac{a}{n}\right) \frac{d\eta}{d\xi} - bc\eta = 0, \qquad \xi = \frac{1}{n^2} x^n, \qquad y = \frac{n}{b} \frac{d\eta}{d\xi} \frac{\xi}{\eta}. \tag{43'}$$

A comparison of this with (39) shows that the solution is

$$\eta = AI_{-\frac{a}{n}}(-bc\xi) + BI_{\frac{a}{n}}(-bc\xi) \cdot (-bc\xi)^{\frac{a}{n}},$$

which in terms of Bessel functions J becomes, by (40),

$$\eta = \xi^{\frac{a}{2n}} \left[AJ_{\frac{a}{n}}(2\sqrt{-bc\xi}) + BJ_{-\frac{a}{n}}(2\sqrt{-bc\xi})\right].$$

* If n is not integral, both $n!$ and $(n+i)!$ must be replaced (§ 147) by $\Gamma(n+1)$ and $\Gamma(n+i+1)$.

The value of y may be found by substitution and use of (29).

$$y = \sqrt{-\frac{c}{b}}\, x^{\frac{n}{2}} \frac{J_{\frac{a}{n}-1}(2x^{\frac{n}{2}}\sqrt{-bc}/n) - AJ_{1-\frac{a}{n}}(2x^{\frac{n}{2}}\sqrt{-bc}/n)}{J_{\frac{a}{n}}(2x^{\frac{n}{2}}\sqrt{-bc}/n) + AJ_{-\frac{a}{n}}(2x^{\frac{n}{2}}\sqrt{-bc}/n)}, \tag{44}$$

where A denotes the one arbitrary constant of integration.

It is noteworthy that the cylinder functions are sometimes expressible in terms of trigonometric functions. For when $n = \frac{1}{2}$ the equation (35) has the integrals

$$y = A\sin x + B\cos x \quad \text{and} \quad y = x^{\frac{1}{2}}[AC_{\frac{1}{2}}(x) + BC_{-\frac{1}{2}}(x)].$$

Hence it is permissible to write the relations

$$x^{\frac{1}{2}}C_{\frac{1}{2}}(x) = \sin x, \qquad x^{\frac{1}{2}}C_{-\frac{1}{2}}(x) = \cos x, \tag{45}$$

where C is a suitably chosen cylinder function of order $\frac{1}{2}$. From these equations by application of (29) the cylinder functions of order $p + \frac{1}{2}$, where p is any integer, may be found.

Now if Riccati's equation is such that b and c have opposite signs and a/n is of the form $p + \frac{1}{2}$, the integral (44) can be expressed in terms of trigonometric functions by using the values of the functions $C_{p+\frac{1}{2}}$ just found in place of the J's. Moreover if b and c have the same sign, the trigonometric solution will still hold formally and may be converted into exponential or hyperbolic form. Thus Riccati's equation is integrable in terms of the elementary functions when $a/n = p + \frac{1}{2}$ no matter what the sign of bc is.

EXERCISES

1. Prove the following relations:

(α) $4C_n'' = C_{n-2} - 2C_n + C_{n+2}$, $\qquad(\beta)$ $xC_n = 2(n+1)C_{n+1} - xC_{n+2}$,

(γ) $2^3C_n''' = C_{n-3} - 3C_{n-1} + 3C_{n+1} - C_{n+3}$, $\qquad$ generalize,

(δ) $xC_n = 2(n+1)C_{n+1} - 2(n+3)C_{n+3} + 2(n+5)C_{n+5} - xC_{n+6}$.

2. Study the functions defined by the pair of relations

$$F_{n-1}(x) + F_{n+1}(x) = 2\frac{d}{dx}F_n(x), \qquad F_{n-1}(x) - F_{n+1}(x) = \frac{2}{x}F_n(x)$$

especially to find results analogous to (30)–(35).

3. Use Ex. 12, p. 247, to obtain (34) and the corresponding relation in Ex. 2.

4. Show that the solution of (38) is $y = AI_n\int\frac{dx}{x^{n+1}I_n^2} + BI_n$.

5. Write out five terms in the expansions of $I_0, I_1, I_{-\frac{1}{2}}, J_0, J_1$.

6. Show from the expansion (42) that $\frac{1}{2}!\sqrt{\frac{2}{x}}J_{\frac{1}{2}}(x) = \frac{1}{x}\sin x$.

7. From (45), (29) obtain the following:

$$x^{\frac{1}{2}}C_{\frac{3}{2}}(x) = \frac{\sin x}{x} - \cos x, \qquad x^{\frac{1}{2}}C_{\frac{5}{2}}(x) = \left(\frac{3}{x^2} - 1\right)\sin x - \frac{3}{x}\cos x,$$

$$x^{\frac{1}{2}}C_{-\frac{3}{2}}(x) = -\sin x - \frac{\cos x}{x}, \qquad x^{\frac{1}{2}}C_{-\frac{5}{2}}(x) = \frac{3}{x}\sin x + \left(\frac{3}{x^2} - 1\right)\cos x.$$

8. Prove by integration by parts: $\int \frac{J_2 dx}{x^3} dx = \frac{J_3}{x^3} + 6\frac{J_4}{x^4} + 6 \cdot 8 \int \frac{J_5 dx}{x^5}$.

9. Suppose $C_n(x)$ and $K_n(x)$ so chosen that $A = 1$ in (34). Show that

$$y = AC_n(x) + BK_n(x) + L\left[K_n(x)\int \frac{C_n(x)}{x^3}dx - C_n(x)\int \frac{K_n(x)}{x^3}dx\right]$$

is the integral of the differential equation $x^2y'' + xy' + (x^2 - n^2)y = Lx^{-2}$.

10. Note that the solution of Riccati's equation has the form

$$y = \frac{f(x) + Ag(x)}{F(x) + AG(x)}, \quad \text{and show that} \quad \frac{dy}{dx} + P(x)y + Q(x)y^2 = R(x)$$

will be the form of the equation which has such an expression for its integral.

11. Integrate these equations in terms of cylinder functions and reduce the results whenever possible by means of Ex. 7:

(α) $xy' - 5y + y^2 + x^2 = 0$, $\qquad (\beta)$ $xy' - 3y + y^2 = x^2$.

(γ) $y'' + ye^{2x} = 0$, $\qquad (\delta)$ $x^2y'' + nxy' + (b + cx^{2m})y = 0$.

12. Identify the functions of Ex. 2 with the cylinder functions of ix.

13. Let $(x^2 - 1)P_n' = (n+1)(P_{n+1} - xP_n), \qquad P_{n+1}' = xP_n' + (n+1)P_n \qquad (46)$

be taken as defining the *Legendre functions* $P_n(x)$ of order n. Prove

(α) $(x^2 - 1)P_n' = n(xP_n - P_{n-1})$, $\qquad (\beta)$ $(2n+1)xP_n = (n+1)P_{n+1} + nP_{n-1}$,

(γ) $(2n+1)P_n = P_{n+1}' - P_{n-1}'$, $\qquad (\delta)$ $(1 - x^2)P_n'' - 2xP_n' + n(n+1)P_n = 0$.

14. Show that $P_nQ_n' - P_n'Q_n = \frac{A}{x^2 - 1}$ and $P_nQ_{n+1} - P_{n+1}Q_n = -\frac{A}{n+1}$,

where P and Q are any two Legendre functions. Express the general solution of the differential equation of Ex. 13 (δ) analogously to Ex. 4.

15. Let $u = x^2 - 1$ and let D denote differentiation by x. Show

$$D^{n+1}u^{n+1} = D^{n+1}(uu^n) = uD^{n+1}u^n + 2(n+1)xD^nu^n + n(n+1)D^{n-1}u^n,$$
$$D^{n+1}u^{n+1} = D^nDu^{n+1} = 2(n+1)D^n(xu^n) = 2(n+1)xD^nu^n + 2n(n+1)D^{n-1}u^n.$$

Hence show that the derivative of the second equation and the eliminant of $D^{n-1}u^n$ between the two equations give two equations which reduce to (46) if

$$P_n(x) = \frac{1}{2^n \cdot n!}\frac{d^n}{dx^n}(x^2 - 1)^n. \qquad \left\{\begin{array}{l}\text{When } n \text{ is integral these are}\\ \textit{Legendre's polynomials.}\end{array}\right.$$

16. Determine the solutions of Ex. 13 (δ) in series for the initial conditions

(α) $P_n(0) = 1$, $P_n'(0) = 0$, $\qquad (\beta)$ $P_n(0) = 0$, $P_n'(0) = 1$.

17. Take $P_0 = 1$ and $P_1 = x$. Show that these are solutions of (46) and compute P_2, P_3, P_4 from Ex. 13 (β). If $x = \cos\theta$, show

$P_2 = \frac{3}{4}\cos 2\theta + \frac{1}{4}$, $\qquad P_3 = \frac{5}{8}\cos 3\theta + \frac{3}{8}\cos\theta$, $\qquad P_4 = \frac{35}{64}\cos 4\theta + \frac{20}{64}\cos 2\theta + \frac{9}{64}$.

18. Write Ex. 13 (δ) as $\frac{d}{dx}[(1 - x^2)P_n'] + n(n+1)P_n = 0$ and show

$$[m(m+1) - n(n+1)]\int_{-1}^{+1} P_nP_m dx = \int_{-1}^{+1}\left[P_m\frac{d(1 - x^2)P_n'}{dx} - P_n\frac{d(1 - x^2)P_m'}{dx}\right]dx.$$

Integrate by parts, assume the functions and their derivatives are finite, and show

$$\int_{-1}^{+1} P_n P_m dx = 0, \quad \text{if} \quad n \neq m.$$

19. By successive integration by parts and by reduction formulas show

$$\int_{-1}^{+1} P_n^2 dx = \frac{1}{2^{2n}(n\,!)^2}\int_{-1}^{+1} \frac{d^n(x^2-1)^n}{dx^n}\cdot\frac{d^n(x^2-1)^n}{dx^n}dx = \frac{(-1)^n}{2^n \cdot n\,!}\int_{-1}^{+1}(x^2-1)^n dx$$

and
$$\int_{-1}^{+1} P_n^2 dx = \frac{2}{2n+1}, \qquad n \text{ integral.}$$

20. Show
$$\int_{-1}^{+1} x^m P_n dx = \int_{-1}^{+1} x^m \frac{d^n(x^2-1)^n}{dx^n} = 0, \qquad \text{if } m < n.$$

Determine the value of the integral when $m = n$. Cannot the results of Exs. 18, 19 for m and n integral be obtained simply from these results?

21. Consider (38) and its solution $I_0 = 1 - x + \frac{x^2}{2\,!^2} - \frac{x^3}{3\,!^2} + \frac{x^4}{4\,!^2} - \cdots$ when $n = 0$. Assume a solution of the form $y = I_0 v + w$ so that

$$x\frac{d^2w}{dx^2} + \frac{dw}{dx} + w + 2x\frac{dI_0}{dx}\frac{dv}{dx} = 0, \quad \text{if} \quad x\frac{d^2v}{dx^2} + \frac{dv}{dx} = 0,$$

is the equation for w if v satisfies the equation $xv'' + v' = 0$. Show

$$v = A + B\log x, \qquad xw'' + w' + w = 2B - \frac{2Bx}{2\,!} + \frac{2Bx^2}{2\,!\,3\,!} - \frac{2Bx^3}{3\,!\,4\,!} + \cdots.$$

By assuming $w = a_1 x + a_2 x^2 + \cdots$, determine the a's and hence obtain

$$w = 2B\left[x - \frac{x^2}{2\,!^2}\left(1 + \frac{1}{2}\right) + \frac{x^3}{3\,!^2}\left(1 + \frac{1}{2} + \frac{1}{3}\right) - \frac{x^4}{4\,!^2}\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) + \cdots\right];$$

and $(A + B\log x)\,I_0 + w$ is then the complete solution containing two constants. As AI_0 is one solution, $B\log x \cdot I_0 + w$ is another. From this second solution for $n = 0$, the second solution for any integral value of n may be obtained by differentiation; the work, however, is long and the result is somewhat complicated.

CHAPTER X

DIFFERENTIAL EQUATIONS IN MORE THAN TWO VARIABLES

109. Total differential equations. An equation of the form

$$P(x, y, z)\,dx + Q(x, y, z)\,dy + R(x, y, z)\,dz = 0, \tag{1}$$

involving the differentials of three variables is called a *total differential equation.* A similar equation in any number of variables would also be called total; but the discussion here will be restricted to the case of three. If definite values be assigned to x, y, z, say a, b, c, the equation becomes

$$A\,dx + B\,dy + C\,dz = A(x - a) + B(y - b) + C(z - c) = 0, \tag{2}$$

where x, y, z are supposed to be restricted to values near a, b, c, and represents a small portion of a plane passing through (a, b, c). From the analogy to the lineal element (§ 85), such a portion of a plane may be called a *planar element.* The differential equation therefore represents an infinite number of planar elements, one passing through each point of space.

Now any family of surfaces $F(x, y, z) = C$ also represents an infinity of planar elements, namely, the portions of the tangent planes at every point of all the surfaces in the neighborhood of their respective points of tangency. In fact

$$dF = F'_x dx + F'_y dy + F'_z dz = 0 \tag{3}$$

is an equation similar to (1). If the planar elements represented by (1) and (3) are to be the same, the equations cannot differ by more than a factor $\mu(x, y, z)$. Hence

$$F'_x = \mu P, \qquad F'_y = \mu Q, \qquad F'_z = \mu R.$$

If a function $F(x, y, z) = C$ can be found which satisfies these conditions, it is said to be the integral of (1), and the factor $\mu(x, y, z)$ by which the equations (1) and (3) differ is called an *integrating factor* of (1). Compare § 91.

It may happen that $\mu = 1$ and that (1) is thus an *exact* differential. In this case the conditions

$$P'_y = Q'_x, \qquad Q'_z = R'_y, \qquad R'_x = P'_z, \tag{4}$$

which arise from $F''_{xy} = F''_{yx}$, $F''_{yz} = F''_{zy}$, $F''_{zx} = F''_{xz}$, must be satisfied. Moreover if these conditions are satisfied, the equation (1) will be an exact equation and the integral is given by

$$F(x, y, z) = \int_{x_0}^{x} P(x, y, z)\, dx + \int_{y_0}^{y} Q(x_0, y, z)\, dy + \int R(x_0, y_0, z)\, dz = C,$$

where x_0, y_0, z_0 may be chosen so as to render the integration as simple as possible. The proof of this is so similar to that given in the case of two variables (§ 92) as to be omitted. In many cases which arise in practice the equation, though not exact, may be made so by an obvious integrating factor.

As an example take $zxdy - yzdx + x^2dz = 0$. Here the conditions (4) are not fulfilled but the integrating factor $1/x^2z$ is suggested. Then

$$\frac{xdy - ydx}{x^2} + \frac{dz}{z} = d\left(\frac{y}{x} + \log z\right)$$

is at once perceived to be an exact differential and the integral is $y/x + \log z = C$. It appears therefore that in this simple case neither the renewed application of the conditions (4) nor the general formula for the integral was necessary. It often happens that both the integrating factor and the integral can be recognized at once as above.

If the equation does not suggest an integrating factor, the question arises, Is there any integrating factor? In the case of two variables (§ 94) there always was an integrating factor. In the case of three variables there may be none. For

$$\begin{array}{l|l}
F''_{xy} = P\dfrac{\partial\mu}{\partial y} + \mu\dfrac{\partial P}{\partial y} = F''_{yx} = Q\dfrac{\partial\mu}{\partial x} + \mu\dfrac{\partial Q}{\partial x}, & R, \\
F''_{yz} = Q\dfrac{\partial\mu}{\partial z} + \mu\dfrac{\partial Q}{\partial z} = F''_{zy} = R\dfrac{\partial\mu}{\partial y} + \mu\dfrac{\partial R}{\partial y}, & P, \\
F''_{zx} = R\dfrac{\partial\mu}{\partial x} + \mu\dfrac{\partial R}{\partial x} = F''_{xz} = P\dfrac{\partial\mu}{\partial z} + \mu\dfrac{\partial P}{\partial z}, & Q.
\end{array}$$

If these equations be multiplied by R, P, Q and added and if the result be simplified, the condition

$$P\left(\frac{\partial Q}{\partial z} - \frac{\partial R}{\partial y}\right) + Q\left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right) + R\left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x}\right) = 0 \qquad (5)$$

is found to be imposed on P, Q, R if there is to be an integrating factor. This is called the *condition of integrability*. For it may be shown conversely that if the condition (5) is satisfied, the equation may be integrated.

Suppose an attempt to integrate (1) be made as follows: First assume that one of the variables is constant (naturally, that one which will

make the resulting equation simplest to integrate), say z. Then $Pdx + Qdy = 0$. Now integrate this simplified equation with an integrating factor or otherwise, and let $F(x, y, z) = \phi(z)$ be the integral, where the constant C is taken as a function ϕ of z. Next try to determine ϕ so that the integral $F(x, y, z) = \phi(z)$ will satisfy (1). To do this, differentiate;

$$F'_x dx + F'_y dy + F'_z dz = d\phi.$$

Compare this equation with (1). Then the equations*

$$F'_x = \lambda P, \qquad F'_y = \lambda Q, \qquad (F'_z - \lambda R)\, dz = d\phi$$

must hold. The third equation $(F'_z - \lambda R)\, dz = d\phi$ may be integrated provided the coefficient $S = F'_z - \lambda R$ of dz is a function of z and ϕ, that is, of z and F alone. This is so in case the condition (5) holds. It therefore appears that the integration of the equation (1) for which (5) holds reduces to the succession of two integrations of the type discussed in Chap. VIII.

As an example take $(2x^2 + 2xy + 2xz^2 + 1)\, dx + dy + 2zdz = 0$. The condition

$$(2x^2 + 2xy + 2xz^2 + 1)0 + 1(-4xz) + 2z(2x) = 0$$

of integrability is satisfied. The greatest simplification will be had by making x constant. Then $dy + 2zdz = 0$ and $y + z^2 = \phi(x)$. Compare

$$dy + 2zdz = d\phi \quad \text{and} \quad (2x^2 + 2xy + 2xz^2 + 1)\, dx + dy + 2zdz = 0.$$

Then $\qquad \lambda = 1, \qquad -(2x^2 + 2xy + 2xz^2 + 1)\, dx = d\phi;$

or $\qquad -(2x^2 + 1 + 2x\phi)\, dx = d\phi \quad \text{or} \quad d\phi + 2x\phi dx = -(2x^2 + 1)\, dx.$

This is the linear type with the integrating factor e^{x^2}. Then

$$e^{x^2}(d\phi + 2x\phi dx) = -e^{x^2}(2x^2 + 1)\, dx \quad \text{or} \quad e^{x^2}\phi = -\int e^{x^2}(2x^2 + 1)\, dx + C.$$

Hence $y + z^2 + e^{-x^2}\int e^{x^2}(2x^2 + 1)\, dx = Ce^{-x^2}$ or $e^{x^2}(y + z^2) + \int e^{x^2}(2x^2 + 1)\, dx = C$

is the solution. It may be noted that e^{x^2} is the integrating factor for the original equation:

$$e^{x^2}[(2x^2 + 2xy + 2xz^2 + 1)\, dx + dy + 2zdz] = d\left[e^{x^2}(y + z^2) + \int e^{x^2}(2x^2 + 1)\, dx\right].$$

To complete the proof that the equation (1) is integrable if (5) is satisfied, it is necessary to show that when the condition is satisfied the coefficient $S = F'_z - \lambda R$ is a function of z and F alone. Let it be regarded as a function of x, F, z instead of x, y, z. It is necessary to prove that the derivative of S by x when F and z are constant is zero. By the formulas for change of variable

$$\left(\frac{\partial S}{\partial x}\right)_{y,z} = \left(\frac{\partial S}{\partial x}\right)_{F,z} + \left(\frac{\partial S}{\partial F}\right)\frac{\partial F}{\partial x}, \qquad \left(\frac{\partial S}{\partial y}\right)_{x,z} = \left(\frac{\partial S}{\partial F}\right)_{x,z}\frac{\partial F}{\partial y}.$$

* Here the factor λ is not an integrating factor of (1), but only of the reduced equation $Pdx + Qdy = 0$.

But $F'_x = \lambda P$ and $F'_y = \lambda Q$, and hence $Q\left(\frac{\partial S}{\partial x}\right)_{y,z} - P\left(\frac{\partial S}{\partial y}\right)_{x,z} = Q\left(\frac{\partial S}{\partial x}\right)_{F,z}$.

Now
$$\left(\frac{\partial S}{\partial x}\right)_{y,z} = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial z} - \lambda R\right) = \frac{\partial^2 F}{\partial z \partial x} - \frac{\partial \lambda R}{\partial x} = \frac{\partial \lambda P}{\partial z} - \frac{\partial \lambda R}{\partial x}.$$

Hence
$$\left(\frac{\partial S}{\partial x}\right)_{y,z} = \lambda\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + P\frac{\partial \lambda}{\partial z} - R\frac{\partial \lambda}{\partial x},$$

and
$$\left(\frac{\partial S}{\partial y}\right)_{x,z} = \lambda\left(\frac{\partial Q}{\partial z} - \frac{\partial R}{\partial y}\right) + Q\frac{\partial \lambda}{\partial z} - R\frac{\partial \lambda}{\partial y}.$$

Then
$$Q\left(\frac{\partial S}{\partial x}\right)_{y,z} - P\left(\frac{\partial S}{\partial y}\right)_{x,z} = \lambda\left[Q\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + P\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\right] - R\left[Q\frac{\partial \lambda}{\partial x} - P\frac{\partial \lambda}{\partial y}\right]$$

and
$$Q\left(\frac{\partial S}{\partial x}\right)_{F,z} = \lambda\left[Q\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + P\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + R\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\right] - R\left[\frac{\partial \lambda Q}{\partial x} - \frac{\partial \lambda P}{\partial y}\right],$$

where a term has been added in the first bracket and subtracted in the second. Now as λ is an integrating factor for $Pdx + Qdy$, it follows that $(\lambda Q)'_x = (\lambda P)'_y$; and only the first bracket remains. By the condition of integrability this, too, vanishes and hence S as a function of x, F, z does not contain x but is a function of F and z alone, as was to be proved.

110. It has been seen that if the equation (1) is integrable, there is an integrating factor and the condition (5) is satisfied; also that conversely if the condition is satisfied the equation may be integrated. Geometrically this means that the infinity of planar elements defined by the equation can be grouped upon a family of surfaces $F(x, y, z) = C$ to which they are tangent. If the condition of integrability is not satisfied, the planar elements cannot be thus grouped into surfaces. Nevertheless if a surface $G(x, y, z) = 0$ be given, the planar element of (1) which passes through any point (x_0, y_0, z_0) of the surface will cut the surface $G = 0$ in a certain lineal element of the surface. Thus upon the surface $G(x, y, z) = 0$ there will be an infinity of lineal elements, one through each point, which satisfy the given equation (1). And these elements may be grouped into curves lying upon the surface. If the equation (1) is integrable, these curves will of course be the intersections of the given surface $G = 0$ with the surfaces $F = C$ defined by the integral of (1).

The method of obtaining the curves upon $G(x, y, z) = 0$ which are the integrals of (1), in case (5) does not possess an integral of the form $F(x, y, z) = C$, is as follows. Consider the two equations

$$Pdx + Qdy + Rdz = 0, \qquad G'_x dx + G'_y dy + G'_z dz = 0,$$

of which the first is the given differential equation and the second is the differential equation of the given surface. From these equations

one of the differentials, say dz, may be eliminated, and the corresponding variable z may also be eliminated by substituting its value obtained by solving $G(x, y, z) = 0$. Thus there is obtained a differential equation $Mdx + Ndy = 0$ connecting the other two variables x and y. The integral of this, $F(x, y) = C$, consists of a family of cylinders which cut the given surface $G = 0$ in the curves which satisfy (1).

Consider the equation $ydx + xdy - (x + y + z)\,dz = 0$. This does not satisfy the condition (5) and hence is not completely integrable; but a set of integral curves may be found on any assigned surface. If the surface be the plane $z = x + y$, then

$$ydx + xdy - (x + y + z)\,dz = 0 \quad \text{and} \quad dz = dx + dy$$

give $\quad (x + z)\,dx + (y + z)\,dy = 0 \quad \text{or} \quad (2x + y)\,dx + (2y + x)\,dy = 0$

by eliminating dz and z. The resulting equation is exact. Hence

$$x^2 + xy + y^2 = C \quad \text{and} \quad z = x + y$$

give the curves which satisfy the equation and lie in the plane.

If the equation (1) were integrable, the integral curves may be used to obtain the integral surfaces and thus to accomplish the complete integration of the equation by *Mayer's method.* For suppose that $F(x, y, z) = C$ were the integral surfaces and that $F(x, y, z) = F(0, 0, z_0)$ were that particular surface cutting the z-axis at z_0. The family of planes $y = \lambda x$ through the z-axis would cut the surface in a series of curves which would be integral curves, and the surface could be regarded as generated by these curves as the plane turned about the axis. To reverse these considerations let $y = \lambda x$ and $dy = \lambda dx$; by these relations eliminate dy and y from (1) and thus obtain the differential equation $Mdx + Ndz = 0$ of the intersections of the planes with the solutions of (1). Integrate the equation as $f(x, z, \lambda) = C$ and determine the constant so that $f(x, z, \lambda) = f(0, z_0, \lambda)$. For any value of λ this gives the intersection of $F(x, y, z) = F(0, 0, z_0)$ with $y = \lambda x$. Now if λ be eliminated by the relation $\lambda = y/x$, the result will be the surface

$$f\left(x, z, \frac{y}{x}\right) = f\left(0, z_0, \frac{y}{x}\right), \quad \text{equivalent to} \quad F(x, y, z) = F(0, 0, z_0),$$

which is the integral of (1) and passes through $(0, 0, z_0)$. As z_0 is arbitrary, the solution contains an arbitrary constant and is the general solution.

It is clear that instead of using planes through the z-axis, planes through either of the other axes might have been used, or indeed planes or cylinders through any line parallel to any of the axes. Such modifications are frequently necessary owing to the fact that the substitution $f(0, z_0, \lambda)$ introduces a division by 0 or a log 0 or some other impossibility. For instance consider

$$y^2dx + zdy - ydz = 0, \qquad y = \lambda x, \qquad dy = \lambda dx, \qquad \lambda^2x^2dx + \lambda zdx - \lambda xdz = 0.$$

Then $$\lambda dx + \frac{zdx - xdz}{x^2} = 0, \quad \text{and} \quad \lambda x - \frac{z}{x} = f(x, z, \lambda).$$

But here $f(0, z_0, \lambda)$ is impossible and the solution is illusory. If the planes $(y - 1) = \lambda x$ passing through a line parallel to the z-axis and containing the point $(0, 1, 0)$ had been used, the result would be

$$dy = \lambda dx, \qquad (1 + \lambda x)^2dx + \lambda zdx - (1 + \lambda x)\,dz = 0,$$

or $$dx + \frac{\lambda z dx - (1 + \lambda x)\,dz}{(1 + \lambda x)^2} = 0, \quad \text{and} \quad x - \frac{z}{1 + \lambda x} = f(x, z, \lambda).$$

Hence $$x - \frac{z}{1 + \lambda x} = - z_0 \quad \text{or} \quad x - \frac{z}{y} = - z_0 = C,$$

is the solution. The same result could have been obtained with $x = \lambda z$ or $y = \lambda\,(x - a)$. In the latter case, however, care should be taken to use $f(x, z, \lambda) = f(a, z_0, \lambda)$.

EXERCISES

1. Test these equations for exactness; if exact, integrate; if not exact, find an integrating factor by inspection and integrate:

(α) $(y + z)\,dx + (z + x)\,dy + (x + y)\,dz = 0$, ($\beta$) $y^2dx + zdy - ydz = 0$,
(γ) $xdx + ydy - \sqrt{a^2 - x^2 - y^2}\,dz = 0$, ($\delta$) $2z\,(dx - dy) + (x - y)\,dz = 0$,
(ϵ) $(2x + y^2 + 2xz)\,dx + 2xydy + x^2dz = 0$, ($\zeta$) $zydx = zxdy + y^2dz$,
(η) $x(y - 1)(z - 1)\,dx + y(z - 1)(x - 1)\,dy + z(x - 1)(y - 1)\,dz = 0$.

2. Apply the test of integrability and integrate these:

(α) $(x^2 - y^2 - z^2)\,dx + 2xydy + 2xzdz = 0$,
(β) $(x + y^2 + z^2 + 1)\,dx + 2ydy + 2zdz = 0$,
(γ) $(y + a)^2dx + zdy = (y + a)\,dz$,
(δ) $(1 - x^2 - 2y^2z)\,dz = 2xzdx + 2yz^2dy$,
(ϵ) $x^2dx^2 + y^2dy^2 - z^2dz^2 + 2xydxdy = 0$,
(ζ) $z\,(xdx + ydy + zdz)^2 = (z^2 - x^2 - y^2)(xdx + ydy + zdz)\,dz$.

3. If the equation is homogeneous, the substitution $x = uz$, $y = vz$, frequently shortens the work. Show that if the given equation satisfies the condition of integrability, the new equation will satisfy the corresponding condition in the new variables and may be rendered exact by an obvious integrating factor. Integrate:

(α) $(y^2 + yz)\,dx + (xz + z^2)\,dy + (y^2 - xy)\,dz = 0$,
(β) $(x^2y - y^3 - y^2z)\,dx + (xy^2 - x^2z - x^3)\,dy + (xy^2 + x^2y)\,dz = 0$,
(γ) $(y^2 + yz + z^2)\,dx + (x^2 + xz + z^2)\,dy + (x^2 + xy + y^2)\,dz = 0$.

4. Show that (5) does not hold; integrate subject to the relation imposed:

(α) $ydx + xdy - (x + y + z)\,dz = 0$, $x + y + z = k$ or $y = kx$,
(β) $c\,(xdy + ydy) + \sqrt{1 - a^2x^2 - b^2y^2}\,dz = 0$, $a^2x^2 + b^2y^2 + c^2z^2 = 1$,
(γ) $dz = aydx + bdy$, $y = kx$ or $x^2 + y^2 + z^2 = 1$ or $y = f(x)$.

5. Show that if an equation is integrable, it remains integrable after any change of variables from x, y, z to u, v, w.

6. Apply Mayer's method to sundry of Exs. 2 and 3.

7. Find the conditions of exactness for an equation in four variables and write the formula for the integration. Integrate with or without a factor:

(α) $(2x + y^2 + 2xz)\,dx + 2xydy + x^2dz + du = 0$,
(β) $yzudx + xzudy + xyudz + xyzdu = 0$,
(γ) $(y + z + u)\,dx + (x + z + u)\,dy + (x + y + u)\,dz + (x + y + z)\,du = 0$,
(δ) $u(y + z)\,dx + u(y + z + 1)\,dy + udz - (y + z)du = 0$.

8. If an equation in four variables is integrable, it must be so when any one of the variables is held constant. Hence the four conditions of integrability obtained by writing (5) for each set of three coefficients must hold. Show that the conditions

are satisfied in the following cases. Find the integrals by a generalization of the method in the text by letting one variable be constant and integrating the three remaining terms and determining the constant of integration as a function of the fourth in such a way as to satisfy the equations.

(α) $z(y+z)\,dx + z(u-x)\,dy + y(x-u)\,dz + y(y+z)\,du = 0.$

(β) $uyz\,dx + uzx\log x\,dy + uxy\log x\,dz - x\,du = 0.$

9. Try to extend the method of Mayer to such as the above in Ex. 8.

10. If $G(x, y, z) = a$ and $H(x, y, z) = b$ are two families of surfaces defining a family of curves as their intersections, show that the equation

$$(G'_y H'_z - G'_z H'_y)\,dx + (G'_z H'_x - G'_x H'_z)\,dy + (G'_x H'_y - G'_y H'_x)\,dz = 0$$

is the equation of the planar elements perpendicular to the curves at every point of the curves. Find the conditions on G and H that there shall be a family of surfaces which cut all these curves orthogonally. Determine whether the curves below have orthogonal trajectories (surfaces); and if they have, find the surfaces:

(α) $y = x + a,\ z = x + b.$ (β) $y = ax + 1,\ z = bx.$

(γ) $x^2 + y^2 = a^2,\ z = b.$ (δ) $xy = a,\ xz = b.$

(ϵ) $x^2 + y^2 + z^2 = a^2,\ xy = b.$ (ζ) $x^2 + 2y^2 + 3z^2 = a,\ xy + z = b.$

(η) $\log xy = az,\ x + y + z = b.$ (θ) $y = 2ax + a^2,\ z = 2bx + b^2.$

11. Extend the work of proposition 3, § 94, and Ex. 11, p. 234, to find the normal derivative of the solution of equation (1) and to show that the singular solution may be looked for among the factors of $\mu^{-1} = 0$.

12. If $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ be formed, show that (1) becomes $\mathbf{F}\cdot d\mathbf{r} = 0$. Show that the condition of exactness is $\nabla\times\mathbf{F} = 0$ by expanding $\nabla\times\mathbf{F}$ as the formal vector product of the operator ∇ and the vector $\mathbf{F}$ (see § 78). Show further that the condition of integrability is $\mathbf{F}\cdot(\nabla\times\mathbf{F}) = 0$ by similar formal expansion.

13. In Ex. 10 consider ∇G and ∇H. Show these vectors are normal to the surfaces $G = a$, $H = b$, and hence infer that $(\nabla G)\times(\nabla H)$ is the direction of the intersection. Finally explain why $d\mathbf{r}\cdot(\nabla G\times\nabla H) = 0$ is the differential equation of the orthogonal family if there be such a family. Show that this vector form of the family reduces to the form above given.

111. Systems of simultaneous equations.

The two equations

$$\frac{dy}{dx} = f(x, y, z), \qquad \frac{dz}{dx} = g(x, y, z) \tag{6}$$

in the two dependent variables y and z and the independent variable x constitute a set of simultaneous equations of the first order. It is more customary to write these equations in the form

$$\frac{dx}{X(x, y, z)} = \frac{dy}{Y(x, y, z)} = \frac{dz}{Z(x, y, z)}, \tag{7}$$

which is symmetric in the differentials and where $X : Y : Z = 1 : f : g$. At any assigned point x_0, y_0, z_0 of space the ratios $dx : dy : dz$ of the differentials are determined by substitution in (7). Hence the equations

fix a definite direction at each point of space, that is, they determine a lineal element through each point. The problem of integration is to combine these lineal elements into a family of curves $F(x, y, z) = C_1$, $G(x, y, z) = C_2$, depending on two parameters C_1 and C_2, one curve passing through each point of space and having at that point the direction determined by the equations.

For the formal integration there are several allied methods of procedure. In the first place it may happen that two of

$$\frac{dx}{X} = \frac{dy}{Y}, \qquad \frac{dy}{Y} = \frac{dz}{Z}, \qquad \frac{dx}{X} = \frac{dz}{Z}$$

are of such a form as to contain only the variables whose differentials enter. In this case these two may be integrated and the two solutions taken together give the family of curves. Or it may happen that one and only one of these equations can be integrated. Let it be the first and suppose that $F(x, y) = C_1$ is the integral. By means of this integral the variable x may be eliminated from the second of the equations or the variable y from the third. In the respective cases there arises an equation which may be integrated in the form $G(y, z, C_1) = C_2$ or $G(x, z, F) = C_2$, and this result taken with $F(x, y) = C_1$ will determine the family of curves.

Consider the example $\dfrac{x\,dx}{yz} = \dfrac{y\,dy}{xz} = \dfrac{dz}{y}$. Here the two equations

$$\frac{x\,dx}{y} = \frac{y\,dy}{x} \quad \text{and} \quad \frac{x\,dx}{z} = dz$$

are integrable with the results $x^3 - y^3 = C_1$, $x^2 - z^2 = C_2$, and these two integrals constitute the solution. The solution might, of course, appear in very different form; for there are an indefinite number of pairs of equations $F(x, y, z, C_1) = 0$, $G(x, y, z, C_2) = 0$ which will intersect in the curves of intersection of $x^3 - y^3 = C_1$, and $x^2 - z^2 = C_2$. In fact $(y^3 + C_1)^2 = (z^2 + C_2)^3$ is clearly a solution and could replace either of those found above.

Consider the example $\dfrac{dx}{x^2 - y^2 - z^2} = \dfrac{dy}{2xy} = \dfrac{dz}{2xz}$. Here

$$\frac{dy}{y} = \frac{dz}{z}, \quad \text{with the integral} \quad y = C_1 z,$$

is the only equation the integral of which can be obtained directly. If y be eliminated by means of this first integral, there results the equation

$$\frac{dx}{x^2 - (C_1^2 + 1)z^2} = \frac{dz}{2xz} \quad \text{or} \quad 2xz\,dx + \left[(C_1^2 + 1)z^2 - x^2\right]dz = 0.$$

This is homogeneous and may be integrated with a factor to give

$$x^2 + (C_1^2 + 1)z^2 = C_2 z \quad \text{or} \quad x^2 + y^2 + z^2 = C_2 z.$$

Hence
$$y = C_1 z, \qquad x^2 + y^2 + z^2 = C_2 z$$

is the solution, and represents a certain family of circles.

Another method of attack is to use composition and division.

$$\frac{dx}{X}=\frac{dy}{Y}=\frac{dz}{Z}=\frac{\lambda dx+\mu dy+\nu dz}{\lambda X+\mu Y+\nu Z}. \tag{8}$$

Here λ, μ, ν may be chosen as any functions of (x, y, z). It may be possible so to choose them that the last expression, taken with one of the first three, gives an equation which may be integrated. With this first integral a second may be obtained as before. Or it may be that two different choices of λ, μ, ν can be made so as to give the two desired integrals. Or it may be possible so to select two sets of multipliers that the equation obtained by setting the two expressions equal may be solved for a first integral. Or it may be possible to choose λ, μ, ν so that the denominator $\lambda X+\mu Y+\nu Z=0$, and so that the numerator (which must vanish if the denominator does) shall give an equation

$$\lambda dx+\mu dy+\nu dz=0 \tag{9}$$

which satisfies the condition (5) of integrability and may be integrated by the methods of § 109.

Consider the equations $\dfrac{dx}{x^2+y^2+yz}=\dfrac{dy}{x^2+y^2-xz}=\dfrac{dz}{(x+y)z}$. Here take λ, μ, ν as $1, -1, -1$; then $\lambda X+\mu Y+\nu Z=0$ and $dx-dy-dz=0$ is integrable as $x-y-z=C_1$. This may be used to obtain another integral. But another choice of λ, μ, ν as x, y, 0, combined with the last expression, gives

$$\frac{xdx+ydx}{(x^2+y^2)(x+y)}=\frac{dz}{(x+y)z} \quad \text{or} \quad \log(x^2+y^2)=\log z^2+C_2.$$

Hence
$$x-y-z=C_1 \quad \text{and} \quad x^2+y^2=C_2z^2$$

will serve as solutions. This is shorter than the method of elimination.

It will be noted that these equations just solved are homogeneous. The substitution $x=uz$, $y=vz$ might be tried. Then

$$\frac{udz+zdu}{u^2+v^2+v}=\frac{vdz+zdv}{u^2+v^2-u}=\frac{dz}{u+v}=\frac{zdu}{v^2-uv+v}=\frac{zdv}{u^2-uv-u},$$

or
$$\frac{du}{v^2-uv+v}=\frac{dv}{u^2-uv-u}=\frac{dz}{z}.$$

Now the first equations do not contain z and may be solved. This always happens in the homogeneous case and may be employed if no shorter method suggests itself.

It need hardly be mentioned that all these methods apply equally to the case where there are more than three equations. The geometric picture, however, fails, although the geometric language may be continued if one wishes to deal with higher dimensions than three. In some cases the introduction of a fourth variable, as

$$\frac{dx}{X}=\frac{dy}{Y}=\frac{dz}{Z}=\frac{dt}{1} \quad \text{or} \quad =\frac{dt}{t}, \tag{10}$$

is useful in solving a set of equations which originally contained only three variables. This is particularly true when X, Y, Z are linear with constant coefficients, in which case the methods of § 98 may be applied with t as independent variable.

112. Simultaneous differential equations of higher order, as

$$\frac{d^2x}{dt^2} = X\left(x, y, \frac{dx}{dt}, \frac{dy}{dt}\right), \qquad \frac{d^2y}{dt^2} = Y\left(x, y, \frac{dx}{dt}, \frac{dy}{dt}\right),$$

$$\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = R\left(r, \phi, \frac{dr}{dt}, \frac{d\phi}{dt}\right), \qquad \frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \Phi\left(r, \phi, \frac{dr}{dt}, \frac{d\phi}{dt}\right),$$

especially those of the second order like these, are of constant occurrence in mechanics; for the acceleration requires second derivatives with respect to the time for its expression, and the forces are expressed in terms of the coördinates and velocities. The complete integration of such equations requires the expression of the dependent variables as functions of the independent variable, generally the time, with a number of constants of integration equal to the sum of the orders of the equations. Frequently even when the complete integrals cannot be found, it is possible to carry out some integrations and replace the given system of equations by fewer equations or equations of lower order containing some constants of integration.

No special or general rules will be laid down for the integration of systems of higher order. In each case some particular combinations of the equations may suggest themselves which will enable an integration to be performed.* In problems in mechanics the principles of energy, momentum, and moment of momentum frequently suggest combinations leading to integrations. Thus if

$$x'' = X, \qquad y'' = Y, \qquad z'' = Z,$$

where accents denote differentiation with respect to the time, be multiplied by dx, dy, dz and added, the result

$$x''dx + y''dy + z''dz = Xdx + Ydy + Zdz \tag{11}$$

contains an exact differential on the left; then if the expression on the right is an exact differential, the integration

$$\tfrac{1}{2}(x'^2 + y'^2 + z'^2) = \int Xdx + Ydy + Zdz + C \tag{11'}$$

* It is possible to differentiate the given equations repeatedly and eliminate all the dependent variables except one. The resulting differential equation, say in x and t, may then be treated by the methods of previous chapters; but this is rarely successful except when the equation is linear.

can be performed. This is *the principle of energy* in its simplest form If two of the equations are multiplied by the chief variable of the other and subtracted, the result is

$$yx'' - xy'' = yX - xY \tag{12}$$

and the expression on the left is again an exact differential; if the right-hand side reduces to a constant or a function of t, then

$$yx' - xy' = \int f(t) + C \tag{12'}$$

is an integral of the equations. This is *the principle of moment of momentum.* If the equations can be multiplied by constants as

$$lx'' + my'' + nz'' = lX + mY + nZ, \tag{13}$$

so that the expression on the right reduces to a function of t, an integration may be performed. This is *the principle of momentum.* These three are the most commonly usable devices.

As an example: Let a particle move in a plane subject to forces attracting it toward the axes by an amount proportional to the mass and to the distance from the axes; discuss the motion. Here the equations of motion are merely

$$m\frac{d^2x}{dt^2} = -kmx, \qquad m\frac{d^2y}{dt^2} = -kmy \quad \text{or} \quad \frac{d^2x}{dt^2} = -kx, \qquad \frac{d^2y}{dt^2} = -ky.$$

Then $\quad dx\dfrac{d^2x}{dt^2} + dy\dfrac{d^2y}{dt^2} = -k(xdx + ydy) \quad$ and $\quad \left(\dfrac{dx}{dt}\right)^2 + \left(\dfrac{dy}{dt}\right)^2 = -k(x^2 + y^2) + C.$

Also $\qquad y\dfrac{d^2x}{dt^2} - x\dfrac{d^2y}{dt^2} = 0 \quad$ and $\quad y\dfrac{dx}{dt} - x\dfrac{dy}{dt} = C'.$

In this case the two principles of energy and moment of momentum give two integrals and the equations are reduced to two of the first order. But as it happens, the original equations could be integrated directly as

$$\frac{d^2x}{dt^2}dx = -kxdx, \qquad \left(\frac{dx}{dt}\right)^2 = -kx^2 + C^2, \qquad \frac{dx}{\sqrt{C^2 - kx^2}} = dt$$

$$\frac{d^2y}{dt^2}dy = -kydy, \qquad \left(\frac{dy}{dt}\right)^2 = -ky^2 + K^2, \qquad \frac{dy}{\sqrt{K^2 - ky^2}} = dt.$$

The constants C^2 and K^2 of integration have been written as squares because they are necessarily positive. The complete integration gives

$$\sqrt{k}x = C\sin(\sqrt{k}t + C_1), \qquad \sqrt{k}y = K\sin(\sqrt{k}t + K_2).$$

As another example: A particle, attracted toward a point by a force equal to $r/m^2 + h^2/r^3$ per unit mass, where m is the mass and h is the double areal velocity and r is the distance from the point, is projected perpendicularly to the radius vector at the distance $\sqrt{mh}$; discuss the motion. In polar coördinates the equations of motion are

$$m\left[\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2\right] = R = -\frac{mr}{m^2} - \frac{mh^2}{r^3}, \qquad \frac{m}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \Phi = 0.$$

The second integrates directly as $r^2 d\phi/dt = h$ where the constant of integration h is twice the areal velocity. Now substitute in the first to eliminate ϕ.

$$\frac{d^2r}{dt^2} - \frac{h^2}{r^3} = -\frac{r}{m^2} - \frac{h^2}{r^3} \quad \text{or} \quad \frac{d^2r}{dt^2} = -\frac{r}{m^2} \quad \text{or} \quad \left(\frac{dr}{dt}\right)^2 = -\frac{r^2}{m^2} + C.$$

Now as the particle is projected perpendicularly to the radius, $dr/dt = 0$ at the start when $r = \sqrt{mh}$. Hence the constant C is h/m. Then

$$\frac{dr}{\sqrt{\frac{h}{m} - \frac{r^2}{m^2}}} = dt \quad \text{and} \quad \frac{r^2 d\phi}{h} = dt \quad \text{give} \quad \frac{\sqrt{mh}\,dr}{r^2\sqrt{1 - \frac{r^2}{hm}}} = d\phi.$$

Hence $$\sqrt{mh}\sqrt{\frac{1}{r^2} - \frac{1}{h}} = \phi + C \quad \text{or} \quad \frac{1}{r^2} - \frac{1}{hm} = \frac{(\phi + C)^2}{mh}.$$

Now if it be assumed that $\phi = 0$ at the start when $r = \sqrt{mh}$, we find $C = 0$.

Hence $$r^2 = \frac{mh}{1 + \phi^2} \quad \text{is the orbit}$$

To find the relation between ϕ and the time,

$$r^2 d\phi = h dt \quad \text{or} \quad \frac{m d\phi}{1 + \phi^2} = dt \quad \text{or} \quad t = m \tan^{-1}\phi,$$

if the time be taken as $t = 0$ when $\phi = 0$. Thus the orbit is found, the expression of ϕ as a function of the time is found, and the expression of r as a function of the time is obtainable. The problem is completely solved. It will be noted that the constants of integration have been determined after each integration by the initial conditions. This simplifies the subsequent integrations which might in fact be impossible in terms of elementary functions without this simplification.

EXERCISES

1. Integrate these equations:

(α) $\dfrac{dx}{yz} = \dfrac{dy}{xz} = \dfrac{dz}{xy}$, (β) $\dfrac{dx}{y^2} = \dfrac{dy}{x^2} = \dfrac{dz}{x^2y^2z^2}$,

(γ) $\dfrac{dx}{xz} = \dfrac{dy}{yz} = \dfrac{dz}{xy}$, (δ) $\dfrac{dx}{yz} = \dfrac{dy}{xz} = \dfrac{dz}{x + y}$,

(ϵ) $-\dfrac{dx}{y} = \dfrac{dy}{x} = \dfrac{dz}{1 + z^2}$, (ζ) $\dfrac{dx}{-1} = \dfrac{dy}{3y + 4z} = \dfrac{dz}{2y + 5z}$.

2. Integrate the equations: (α) $\dfrac{dx}{bz - cy} = \dfrac{dy}{cx - az} = \dfrac{dz}{ay - bx}$,

(β) $\dfrac{dx}{x^2 + y^2} = \dfrac{dy}{2xy} = \dfrac{dz}{xz + yz}$, (γ) $\dfrac{dx}{y + z} = \dfrac{dy}{x + z} = \dfrac{dz}{x + y}$,

(δ) $\dfrac{dx}{y^3x - 2x^4} = \dfrac{dy}{2y^4 - x^3y} = \dfrac{dz}{z(x^3 - y^3)}$, (ϵ) $\dfrac{dx}{x(y - z)} = \dfrac{dy}{y(z - x)} = \dfrac{dz}{z(x - y)}$,

(ζ) $\dfrac{dx}{x(y^2 - z^2)} = \dfrac{dy}{y(z^2 - x^2)} = \dfrac{dz}{z(x^2 - y^2)}$, (η) $\dfrac{dx}{x(y^2 - z^2)} = \dfrac{-dy}{y(z^2 + x^2)} = \dfrac{dz}{z(x^2 + y^2)}$,

(θ) $\dfrac{dx}{y - z} = \dfrac{dy}{x + y} = \dfrac{dz}{x + z} = dt$. (ι) $\dfrac{dx}{y - z} = \dfrac{dy}{x + y + t} = \dfrac{dz}{x + z + t} = dt$.

3. Show that the differential equations of the orthogonal trajectories (curves of the family of surfaces $F(x, y, z) = C$ are $dx : dy : dz = F'_x : F'_y : F'_z$. Find the curves which cut the following families of surfaces orthogonally:

(α) $a^2x^2 + b^2y^2 + c^2z^2 = C$, (β) $xyz = C$, (γ) $y^2 = Cxz$,

(δ) $y = x \tan(z + C)$, (ϵ) $y = x \tan Cz$, (ζ) $z = Cxy$.

4. Show that the solution of $dx : dy : dz = X : Y : Z$, where X, Y, Z are linear expressions in x, y, z, can always be found provided a certain cubic equation can be solved.

5. Show that the solutions of the two equations

$$\frac{dx}{dt} + T(ax + by) = T_1, \qquad \frac{dy}{dt} + T(a'x + b'y) = T_2,$$

where T, T_1, T_2 are functions of t, may be obtained by adding the equation as

$$\frac{d}{dt}(x + ly) + \lambda T(x + ly) = T_1 + lT_2$$

after multiplying one by l, and by determining λ as a root of

$$\lambda^2 - (a + b')\lambda + ab' - a'b = 0.$$

6. Solve: (α) $t\dfrac{dx}{dt} + 2(x - y) = t, \qquad t\dfrac{dy}{dt} + x + 5y = t^2.$

(β) $tdx = (t - 2x)\,dt, \qquad tdy = (tx + ty + 2x - t)\,dt,$

(γ) $\dfrac{ldx}{mn(y - z)} = \dfrac{mdy}{nl(z - x)} = \dfrac{ndz}{lm(x - y)} = \dfrac{dt}{t}.$

7. A particle moves in vacuo in a vertical plane under the force of gravity alone. Integrate. Determine the constants if the particle starts from the origin with a velocity V and at an angle of α degrees with the horizontal and at the time $t = 0$.

8. Same problem as in Ex. 7 except that the particle moves in a medium which resists proportionately to the velocity of the particle.

9. A particle moves in a plane about a center of force which attracts proportionally to the distance from the center and to the mass of the particle.

10. Same as Ex. 9 but with a repulsive force instead of an attracting force.

11. A particle is projected parallel to a line toward which it is attracted with a force proportional to the distance from the line.

12. Same as Ex. 11 except that the force is inversely proportional to the square of the distance and only the path of the particle is wanted.

13. A particle is attracted toward a center by a force proportional to the square of the distance. Find the orbit.

14. A particle is placed at a point which repels with a constant force under which the particle moves away to a distance a where it strikes a peg and is deflected off at a right angle with undiminished velocity. Find the orbit of the subsequent motion.

15. Show that equations (7) may be written in the form $d\mathbf{r} \times \mathbf{F} = 0$. Find the condition on $\mathbf{F}$ or on X, Y, Z that the integral curves have orthogonal surfaces.

113. Introduction to partial differential equations. An equation which contains a dependent variable, two or more independent variables, and one or more partial derivatives of the dependent variable with respect to the independent variables is called a *partial differential equation.* The equation

$$P(x, y, z)\frac{\partial z}{\partial x} + Q(x, y, z)\frac{\partial z}{\partial y} = R(x, y, z), \qquad p = \frac{\partial z}{\partial x}, \qquad q = \frac{\partial z}{\partial y}, \quad (14)$$

is clearly a linear partial differential equation of the first order in one dependent and two independent variables. The discussion of this equation preliminary to its integration may be carried on by means of the concept of *planar elements*, and the discussion will immediately suggest the method of integration.

When any point (x_0, y_0, z_0) of space is given, the coefficients P, Q, R in the equation take on definite values and the derivatives p and q are connected by a linear relation. Now any planar element through (x_0, y_0, z_0) may be considered as specified by the two slopes p and q; for it is an infinitesimal portion of the plane $z - z_0 = p(x - x_0) + q(y - y_0)$ in the neighborhood of the point. This plane contains the line or lineal element whose direction is

$$dx : dy : dz = P : Q : R, \quad (15)$$

because the substitution of P, Q, R for $dx = x - x_0$, $dy = y - y_0$, $dz = z - z_0$ in the plane gives the original equation $Pp + Qq = R$. Hence it appears that the planar elements defined by (14), of which there are an infinity through each point of space, are so related that all which pass through a given point of space pass through a certain line through that point, namely the line (15).

Now the problem of integrating the equation (14) is that of grouping the planar elements which satisfy it into surfaces. As at each point they are already grouped in a certain way by the lineal elements through which they pass, it is first advisable to group these lineal elements into curves by integrating the simultaneous equations (15). The integrals of these equations are the curves defined by two families of surfaces $F(x, y, z) = C_1$ and $G(x, y, z) = C_2$. These curves are called the *characteristic curves* or merely the *characteristics* of the equation (14). Through each lineal element of these curves there pass an infinity of the planar elements which satisfy (14). It is therefore clear that if these curves be in any wise grouped into surfaces, the planar elements of the surfaces must satisfy (14); for through each point of the surfaces will pass one of the curves, and the planar element of the surface at that point must therefore pass through the lineal element of the curve and hence satisfy (14).

To group the curves $F(x, y, z) = C_1$, $G(x, y, z) = C_2$ which depend on two parameters C_1, C_2 into a surface, it is merely necessary to introduce some functional relation $C_2 = f(C_1)$ between the parameters so that when one of them, as C_1, is given, the other is determined, and thus a particular curve of the family is fixed by one parameter alone and will sweep out a surface as the parameter varies. Hence *to integrate* (14), *first integrate* (15) *and then write*

$$G(x, y, z) = \Phi[F(x, y, z)] \quad \text{or} \quad \Phi(F, G) = 0, \tag{16}$$

where Φ denotes any arbitrary function. This will be the integral of (14) and will contain an arbitrary function Φ.

As an example, integrate $(y - z)p + (z - x)q = x - y$. Here the equations

$$\frac{dx}{y - z} = \frac{dy}{z - x} = \frac{dz}{x - y} \quad \text{give} \quad x^2 + y^2 + z^2 = C_1, \quad x + y + z = C_2$$

as the two integrals. Hence the solution of the given equation is

$$x + y + z = \Phi(x^2 + y^2 + z^2) \quad \text{or} \quad \Phi(x^2 + y^2 + z^2, x + y + z) = 0,$$

where Φ denotes an arbitrary function. The arbitrary function allows a solution to be determined which shall pass through any desired curve; for if the curve be $f(x, y, z) = 0$, $g(x, y, z) = 0$, the elimination of x, y, z from the four simultaneous equations

$$F(x, y, z) = C_1, \qquad G(x, y, z) = C_2, \qquad f(x, y, z) = 0, \qquad g(x, y, z) = 0$$

will express the condition that the four surfaces meet in a point, that is, that the curve given by the first two will cut that given by the second two; and this elimination will determine a relation between the two parameters C_1 and C_2 which will be precisely the relation to express the fact that the integral curves cut the given curve and that consequently the surface of integral curves passes through the given curve. Thus in the particular case here considered, suppose the solution were to pass through the curve $y = x^2$, $z = x$; then

$$x^2 + y^2 + z^2 = C_1, \qquad x + y + z = C_2, \qquad y = x^2, \qquad z = x$$

give $$2x^2 + x^4 = C_1, \qquad x^2 + 2x = C_2,$$

whence $$(C_2^2 + 2C_2 - C_1)^2 + 8C_2^2 - 24C_1 - 16C_1C_2 = 0.$$

The substitution of $C_1 = x^2 + y^2 + z^2$ and $C_2 = x + y + z$ in this equation will give the solution of $(y - z)p + (z - x)q = x - y$ which passes through the parabola $y = x^2$, $z = x$.

114. It will be recalled that the integral of an ordinary differential equation $f(x, y, y', \cdots, y^{(n)}) = 0$ of the nth order contains n constants, and that conversely if a system of curves in the plane, say $F(x, y, C_1, \cdots, C_n) = 0$, contains n constants, the constants may be eliminated from the equation and its first n derivatives with respect to x. It has now been seen that the integral of a certain partial differential equation contains an arbitrary function, and it might be

inferred that the elimination of an arbitrary function would give rise to a partial differential equation of the first order. To show this, suppose $F(x, y, z) = \Phi[G(x, y, z)]$. Then

$$F'_x + F'_z p = \Phi' \cdot (G'_x + G'_z p), \qquad F'_y + F'_z q = \Phi' \cdot (G'_y + G'_z q)$$

follow from partial differentiation with respect to x and y; and

$$(F'_z G'_y - F'_y G'_z) p + (F'_x G'_z - F'_z G'_x) q = F'_y G'_x - F'_x G'_y$$

is a partial differential equation arising from the elimination of Φ'. More generally, the elimination of n arbitrary functions will give rise to an equation of the nth order; conversely it may be believed that the integration of such an equation would introduce n arbitrary functions in the general solution.

As an example, eliminate from $z = \Phi(xy) + \Psi(x + y)$ the two arbitrary functions Φ and Ψ. The first differentiation gives

$$p = \Phi' \cdot y + \Psi', \qquad q = \Phi' \cdot x + \Psi', \qquad p - q = (y - x)\Phi'.$$

Now differentiate again and let $r = \dfrac{\partial^2 z}{\partial x^2}$, $s = \dfrac{\partial^2 z}{\partial x \partial y}$, $t = \dfrac{\partial^2 z}{\partial y^2}$. Then

$$r - s = -\Phi' + (y - x)\Phi'' \cdot y, \qquad s - t = \Phi' + (y - x)\Phi'' \cdot x.$$

These two equations with $p - q = (y - x)\Phi'$ make three from which

$$xr - (x + y)s + yt = \frac{x + y}{x - y}(p - q) \quad \text{or} \quad x\frac{\partial^2 z}{\partial x^2} - (x + y)\frac{\partial^2 z}{\partial x \partial y} + y\frac{\partial^2 z}{\partial y^2} = \frac{x + y}{x - y}\left(\frac{\partial z}{\partial x} - \frac{\partial z}{\partial y}\right)$$

may be obtained as a partial differential equation of the second order free from Φ and Ψ. The general integral of this equation would be $z = \Phi(xy) + \Psi(x + y)$.

A partial differential equation may represent a certain definite type of surface. For instance by definition a conoidal surface is a surface generated by a line which moves parallel to a given plane, the director plane, and cuts a given line, the directrix. If the director plane be taken as $z = 0$ and the directrix be the z-axis, the equations of any line of the surface are

$$z = C_1, \qquad y = C_2 x, \qquad \text{with} \quad C_1 = \Phi(C_2)$$

as the relation which picks out a definite family of the lines to form a particular conoidal surface. Hence $z = \Phi(y/x)$ may be regarded as the general equation of a conoidal surface of which $z = 0$ is the director plane and the z-axis the directrix. The elimination of Φ gives $px + qy = 0$ as the differential equation of any such conoidal surface.

Partial differentiation may be used not only to eliminate arbitrary functions, but to eliminate constants. For if an equation $f(x, y, z, C_1, C_2) = 0$ contained two constants, the equation and its first derivatives with respect to x and y would yield three equations from which the constants could

be eliminated, leaving a partial differential equation $F(x, y, z, p, q) = 0$ of the first order. If there had been five constants, the equation with its two first derivatives and its three second derivatives with respect to x and y would give a set of six equations from which the constants could be eliminated, leaving a differential equation of the second order. And so on. As the differential equation is obtained by eliminating the constants, the original equation will be a solution of the resulting differential equation.

For example, eliminate from $z = Ax^2 + 2Bxy + Cy^2 + Dx + Ey$ the five constants. The two first and three second derivatives are

$$p = 2Ax + 2By + D, \quad q = 2Bx + 2Cy + E, \quad r = 2A, \quad s = 2B, \quad t = 2C.$$

Hence $$z = -\tfrac{1}{2}rx^2 - \tfrac{1}{2}ty^2 - sxy + px + qy$$

is the differential equation of the family of surfaces. The family of surfaces do not constitute the general solution of the equation, for that would contain two arbitrary functions, but they give what is called a *complete solution*. If there had been only three or four constants, the elimination would have led to a differential equation of the second order which need have contained only one or two of the second derivatives instead of all three; it would also have been possible to find three or two simultaneous partial differential equations by differentiating in different ways.

115. If $$f(x, y, z, C_1, C_2) = 0 \quad \text{and} \quad F(x, y, z, p, q) = 0 \qquad (17)$$

are two equations of which the second is obtained by the elimination of the two constants from the first, the first is said to be the *complete solution* of the second. That is, any equation which contains two distinct arbitrary constants and which satisfies a partial differential equation of the first order is said to be a complete solution of the differential equation. A complete solution has an interesting geometric interpretation. The differential equation $F = 0$ defines a series of planar elements through each point of space. So does $f(x, y, z, C_1, C_2) = 0$. For the tangent plane is given by

$$\left.\frac{\partial f}{\partial x}\right|_0 (x - x_0) + \left.\frac{\partial f}{\partial y}\right|_0 (y - y_0) + \left.\frac{\partial f}{\partial z}\right|_0 (z - z_0) = 0$$

with $$f(x_0, y_0, z_0, C_1, C_2) = 0$$

as the condition that C_1 and C_2 shall be so related that the surface passes through (x_0, y_0, z_0). As there is only this one relation between the two arbitrary constants, there is a whole series of planar elements through the point. As $f(x, y, z, C_1, C_2) = 0$ satisfies the differential equation, the planar elements defined by it are those defined by the differential equation. Thus a complete solution establishes an arrangement of the planar elements defined by the differential equation upon a family of surfaces dependent upon two arbitrary constants of integration.

From the idea of a solution of a partial differential equation of the first order as a surface pieced together from planar elements which satisfy the equation, it appears that the envelope (p. 140) of any family of solutions will itself be a solution; for each point of the envelope is a point of tangency with some one of the solutions of the family, and the planar element of the envelope at that point is identical with the planar element of the solution and hence satisfies the differential equation. *This observation allows the general solution to be determined from any complete solution.* For if in $f(x, y, z, C_1, C_2) = 0$ any relation $C_2 = \Phi(C_1)$ is introduced between the two arbitrary constants, there arises a family depending on one parameter, and the envelope of the family is found by eliminating C_1 from the three equations

$$C_2 = \Phi(C_1), \qquad \frac{\partial f}{\partial C_1} + \frac{d\Phi}{dC_1}\frac{\partial f}{\partial C_2} = 0, \qquad f = 0. \tag{18}$$

As the relation $C_2 = \Phi(C_1)$ contains an arbitrary function Φ, the result of the elimination may be considered as containing an arbitrary function even though it is generally impossible to carry out the elimination except in the case where Φ has been assigned and is therefore no longer arbitrary.

A family of surfaces $f(x, y, z, C_1, C_2) = 0$ depending on two parameters may also have an envelope (p. 139). This is found by eliminating C_1 and C_2 from the three equations

$$f(x, y, z, C_1, C_2) = 0, \qquad \frac{\partial f}{\partial C_1} = 0, \qquad \frac{\partial f}{\partial C_2} = 0.$$

This surface is tangent to all the surfaces in the complete solution. This envelope is called the *singular solution* of the partial differential equation. As in the case of ordinary differential equations (§ 101), the singular solution may be obtained directly from the equation;* it is merely necessary to eliminate p and q from the three equations

$$F(x, y, z, p, q) = 0, \qquad \frac{\partial F}{\partial p} = 0, \qquad \frac{\partial F}{\partial q} = 0.$$

The last two equations express the fact that $F(p, q) = 0$ regarded as a function of p and q should have a double point (§ 57). A reference to § 67 will bring out another point, namely, that not only are all the surfaces represented by the complete solution tangent to the singular solution, but so is any surface which is represented by the general solution.

* It is hardly necessary to point out the fact that, as in the case of ordinary equations, extraneous factors may arise in the elimination, whether of C_1, C_2 or of p, q.

EXERCISES

1. Integrate these linear equations:

(α) $xzp + yzq = xy$, (β) $a(p+q) = z$, (γ) $x^2p + y^2q = z^2$,
(δ) $-yp + xq + 1 + z^2 = 0$, (ϵ) $yp - xq = x^2 - y^2$, (ζ) $(x+z)p = y$,
(η) $x^2p - xyq + y^2 = 0$, (θ) $(a-x)p + (b-y)q = c - z$,
(ι) $p\tan x + q\tan y = \tan z$, (κ) $(y^2 + z^2 - x^2)p - 2xyq + 2xz = 0$.

2. Determine the integrals of the preceding equations to pass through the curves:

for (α) $x^2 + y^2 = 1,\ z = 0$, for (β) $y = 0,\ x = z$,
for (γ) $y = 2x,\ z = 1$, for (ϵ) $x = z,\ y = z$.

3. Show analytically that if $F(x, y, z) = C_1$ is a solution of (15), it is a solution of (14). State precisely what is meant by a solution of a partial differential equation, that is, by the statement that $F(x, y, z) = C_1$ satisfies the equation. Show that the equations

$$P\frac{\partial z}{\partial x} + Q\frac{\partial z}{\partial y} = R \quad \text{and} \quad P\frac{\partial F}{\partial x} + Q\frac{\partial F}{\partial y} + R\frac{\partial F}{\partial z} = 0$$

are equivalent and state what this means. Show that if $F = C_1$ and $G = C_2$ are two solutions, then $F = \Phi(G)$ is a solution, and show conversely that a functional relation must exist between any two solutions (see § 62).

4. Generalize the work in the text along the analytic lines of Ex. 3 to establish the rules for integrating a linear equation in one dependent and four or n independent variables. In particular show that the integral of

$$P_1\frac{\partial z}{\partial x_1} + \cdots + P_n\frac{\partial z}{\partial x_n} = P_{n+1} \quad \text{depends on} \quad \frac{dx_1}{P_1} = \cdots = \frac{dx_n}{P_n} = \frac{dz}{P_{n+1}},$$

and that if $F_1 = C_1, \cdots, F_n = C_n$ are n integrals of the simultaneous system, the integral of the partial differential equation is $\Phi(F_1, \cdots, F_n) = 0$.

5. Integrate: (α) $x\dfrac{\partial u}{\partial x} + y\dfrac{\partial u}{\partial y} + z\dfrac{\partial u}{\partial z} = xyz$,

$$(\beta)\ (y + z + u)\frac{\partial u}{\partial x} + (z + u + x)\frac{\partial u}{\partial y} + (u + x + y)\frac{\partial u}{\partial z} = x + y + z.$$

6. Interpret the general equation of the first order $F(x, y, z, p, q) = 0$ as determining at each point (x_0, y_0, z_0) of space a series of planar elements tangent to a certain cone, namely, the cone found by eliminating p and q from the three simultaneous equations

$$F(x_0, y_0, z_0, p, q) = 0, \qquad (x - x_0)p + (y - y_0)q = z - z_0,$$
$$(x - x_0)\frac{\partial F}{\partial q} - (y - y_0)\frac{\partial F}{\partial p} = 0.$$

7. Eliminate the arbitrary functions:

(α) $x + y + z = \Phi(x^2 + y^2 + z^2)$, (β) $\Phi(x^2 + y^2, z - xy) = 0$,
(γ) $z = \Phi(x + y) + \Psi(x - y)$, (δ) $z = e^{y}\Phi(x - y)$,
(ϵ) $z = y^2 + 2\Phi(x^{-1} + \log y)$, (ζ) $\Phi\left(\dfrac{x}{y}, \dfrac{y}{z}, \dfrac{z}{x}\right) = 0$.

8. Find the differential equations of these types of surfaces:

(α) cylinders with generators parallel to the line $x = az$, $y = bz$,

(β) conical surfaces with vertex at (a, b, c),

(γ) surfaces of revolution about the line $x : y : z = a : b : c$.

9. Eliminate the constants from these equations:

(α) $z = (x + a)(y + b)$, (β) $a(x^2 + y^2) + bz^2 = 1$,

(γ) $(x - a)^2 + (y - b)^2 + (z - c)^2 = 1$, ($\delta$) $(x - a)^2 + (y - b)^2 + (z - c)^2 = d^2$,

(ϵ) $Ax^2 + Bxy + Cy^2 + Dxz + Eyz = z^2$.

10. Show geometrically and analytically that $F(x, y, z) + aG(x, y, z) = b$ is a complete solution of the linear equation.

11. How many constants occur in the complete solution of the equation of the third, fourth, or nth order?

12. Discuss the complete, general, and singular solutions of an equation of the first order $F(x, y, z, u, u'_x, u'_y, u'_z) = 0$ with three independent variables.

13. Show that the planes $z = ax + by + C$, where a and b are connected by the relation $F(a, b) = 0$, are complete solutions of the equation $F(p, q) = 0$. Integrate:

(α) $pq = 1$, (β) $q = p^2 + 1$, (γ) $p^2 + q^2 = m^2$,

(δ) $pq = k$, (ϵ) $k \log q + p = 0$, (ζ) $3p^2 - 2q^2 = 4pq$,

and determine also the singular solutions.

14. Note that a simple change of variable will often reduce an equation to the type of Ex. 13. Thus the equations

$$F\left(\frac{p}{z}, \frac{q}{z}\right) = 0, \qquad F(xp, q) = 0, \qquad F\left(\frac{xp}{z}, \frac{yq}{z}\right) = 0,$$

with $\qquad z = e^{z'}, \qquad x = e^{x'}, \qquad z = e^{z'}, x = e^{x'}, y = e^{y'}$,

take a simpler form. Integrate and determine the singular solutions:

(α) $q = z + px$, (β) $x^2p^2 + y^2q^2 = z^2$, (γ) $z = pq$,

(δ) $q = 2yp^2$, (ϵ) $(p - y)^2 + (q - x)^2 = 1$, ($\zeta$) $z = p^m q^m$.

15. What is the obvious complete solution of the extended Clairaut equation $z = xp + yq + f(p, q)$? Discuss the singular solution. Integrate the equations:

(α) $z = xp + yq + \sqrt{p^2 + q^2 + 1}$, ($\beta$) $z = xp + yq + (p + q)^2$,

(γ) $z = xp + yq + pq$, (δ) $z = xp + yq - 2\sqrt{pq}$.

116. Types of partial differential equations. In addition to the linear equation and the types of Exs. 13–15 above, there are several types which should be mentioned. Of these the first is *the general equation of the first order*. If $F(x, y, z, p, q) = 0$ is the given equation and if a second equation $\Phi(x, y, z, p, q, a) = 0$, which holds simultaneously with the first and contains an arbitrary constant can be found, the two equations may be solved together for the values of p and q, and the results may be substituted in the relation $dz = pdx + qdy$ to give a total differential equation of which the integral will contain the constant a and a second constant of integration b. This integral will then

be a complete integral of the given equation; the general integral may then be obtained by (18) of § 115. This is known as *Charpit's method.*

To find a relation $\Phi = 0$ differentiate the two equations

$$F(x, y, z, p, q) = 0, \qquad \Phi(x, y, z, p, q, a) = 0 \tag{19}$$

with respect to x and y and use the relation that dz be exact.

$$\begin{array}{r|c}
F'_x + F'_z p + F'_p \dfrac{dp}{dx} + F'_q \dfrac{dq}{dx} = 0, & \Phi'_p, \\
\Phi'_x + \Phi'_z p + \Phi'_p \dfrac{dp}{dx} + \Phi'_q \dfrac{dq}{dx} = 0, & -F'_p, \\
F'_y + F'_z q + F'_p \dfrac{dp}{dy} + F'_q \dfrac{dq}{dy} = 0, & \Phi'_q, \\
\Phi'_y + \Phi'_z q + \Phi'_p \dfrac{dp}{dy} + \Phi'_q \dfrac{dq}{dy} = 0, & -F'_q, \\
\dfrac{dp}{dy} - \dfrac{dq}{dx} = 0, & F'_q\Phi'_p - \Phi'_q F'_p.
\end{array}$$

Multiply by the quantities on the right and add. Then

$$(F'_x + pF'_z)\frac{\partial \Phi}{\partial p} + (F'_y + qF'_z)\frac{\partial \Phi}{\partial q} - F'_p\frac{\partial \Phi}{\partial x} - F'_q\frac{\partial \Phi}{\partial y} - (pF'_p + qF'_q)\frac{\partial \Phi}{\partial z} = 0. \tag{20}$$

Now this is a linear equation for Φ and is equivalent to

$$\frac{dp}{F'_x + pF'_z} = \frac{dq}{F'_y + qF'_z} = \frac{dx}{-F'_p} = \frac{dy}{-F'_q} = \frac{dz}{-(pF'_p + qF'_q)} = \frac{d\Phi}{0}. \tag{21}$$

Any integral of this system containing p or q and a will do for Φ, and the simplest integral will naturally be chosen.

As an example take $zp(x+y) + p(q-p) - z^2 = 0$. Then Charpit's equations are

$$\frac{dp}{-zp + p^2(x+y)} = \frac{dq}{zp - 2zq + pq(x+y)} = \frac{dx}{2p - q - z(x+y)}$$
$$= \frac{dy}{-p} = \frac{dz}{2p^2 - 2pq - pz(x+y)}.$$

How to combine these so as to get a solution is not very clear. Suppose the substitution $z = e^{z'}$, $p = e^{z'}p'$, $q = e^{z'}q'$ be made in the equation. Then

$$p'(x+y) + p'(q'-p') - 1 = 0$$

is the new equation. For this Charpit's simultaneous system is

$$\frac{dp'}{p'} = \frac{dq'}{p'} = \frac{dx}{2p' - q' - (x+y)} = \frac{dy}{-p'} = \frac{dz}{2p'^2 - 2pq - p'(x+y)}.$$

The first two equations give at once the solution $dp' = dq'$ or $q' = p' + a$. Solving

$$p'(x+y) + p'(q'-p') - 1 = 0 \quad \text{and} \quad q' = p' + a,$$

$$p' = \frac{1}{a+x+y}, \qquad q' = \frac{1}{a+x+y} + a, \qquad dz' = \frac{dx+dy}{a+x+y} + ady.$$

Then $\quad z = \log(a + x + y) + ay + b \quad$ or $\quad \log z = \log(a + x + y) + ay + b$

is a complete solution of the given equation. This will determine the general integral by eliminating a between the three equations

$$z = e^{ay+b}(a + x + y), \qquad b = f(a), \qquad 0 = (y + f'(a))(a + x + y) + 1,$$

where $f(a)$ denotes an arbitrary function. The rules for determining the singular solution give $z = 0$; but it is clear that the surfaces in the complete solution cannot be tangent to the plane $z = 0$ and hence the result $z = 0$ must be not a singular solution but an extraneous factor. There is no singular solution.

The method of solving a partial differential equation of higher order than the first is to reduce it first to an equation of the first order and then to complete the integration. Frequently the form of the equation will suggest some method easily applied. For instance, if the derivatives of lower order corresponding to one of the independent variables are absent, an integration may be performed as if the equation were an ordinary equation with that variable constant, and the constant of integration may be taken as a function of that variable. Sometimes a change of variable or an interchange of one of the independent variables with the dependent variable will simplify the equation. In general the solver is left mainly to his own devices. Two special methods will be mentioned below.

117. If the equation is *linear with constant coefficients* and all the derivatives are of the same order, the equation is

$$(a_0 D_x^n + a_1 D_x^{n-1} D_y + \cdots + a_{n-1} D_x D_y^{n-1} + a_n D_y^n) z = R(x, y). \qquad (22)$$

Methods like those of § 95 may be applied. Factor the equation.

$$a_0(D_x - \alpha_1 D_y)(D_x - \alpha_2 D_y) \cdots (D_x - \alpha_n D_y) z = R(x, y). \qquad (22')$$

Then the equation is reduced to a succession of equations

$$D_x z - a D_y z = R(x, y),$$

each of which is linear of the first order (and with constant coefficients). Short cuts analogous to those previously given may be developed, but will not be given. If the derivatives are not all of the same order but the polynomial can be factored into linear factors, the same method will apply. For those interested, the several exercises given below will serve as a synopsis for dealing with these types of equation.

There is one equation of the second order,* namely

$$\frac{1}{V^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}, \qquad (23)$$

* This is one of the important differential equations of physics; other important equations and methods of treating them are discussed in Chap. XX.

which occurs constantly in the discussion of waves and which has therefore the name of the *wave equation.* The solution may be written down by inspection. For try the form

$$u(x, y, z, t) = F(ax + by + cz - Vt) + G(ax + by + cz + Vt). \quad (24)$$

Substitution in the equation shows that this is a solution if the relation $a^2 + b^2 + c^2 = 1$ holds, no matter what functions F and G may be. Note that the equation

$$ax + by + cz - Vt = 0, \qquad a^2 + b^2 + c^2 = 1,$$

is the equation of a plane at a perpendicular distance Vt from the origin along the direction whose cosines are a, b, c. If t denotes the time and if the plane moves away from the origin with a velocity V, the function $F(ax + by + cz - Vt) = F(0)$ remains constant; and if $G = 0$, the value of u will remain constant. Thus $u = F$ represents a phenomenon which is constant over a plane and retreats with a velocity V, that is, a plane wave. In a similar manner $u = G$ represents a plane wave approaching the origin. The general solution of (23) therefore represents the superposition of an advancing and a retreating plane wave.

To Monge is due a method sometimes useful in treating differential equations of the second order linear in the derivatives r, s, t; it is known as *Monge's method.* Let

$$Rr + Ss + Tt = V \quad (25)$$

be the equation, where R, S, T, V are functions of the variables and the derivatives p and q. From the given equation and

$$dp = rdx + sdy, \qquad dq = sdx + tdy,$$

the elimination of r and t gives the equation

$$s(Rdy^2 - Sdxdy + Tdx^2) - (Rdydp + Tdxdq - Vdxdy) = 0,$$

and this will surely be satisfied if the two equations

$$Rdy^2 - Sdxdy + Tdx^2 = 0, \qquad Rdydp + Tdxdq - Vdxdy = 0 \quad (25')$$

can be satisfied simultaneously. The first may be factored as

$$dy - f_1(x, y, z, p, q)dx = 0, \qquad dy - f_2(x, y, z, p, q)dx = 0. \quad (26)$$

The problem then is reduced to integrating the system consisting of one of these factors with (25′) and $dz = pdx + qdy$, that is, a system of three total differential equations.

If two independent solutions of this system can be found, as

$$u_1(x, y, z, p, q) = C_1, \qquad u_2(x, y, z, p, q) = C_2,$$

then $u_1 = \Phi(u_2)$ is a first or intermediary integral of the given equation, the general integral of which may be found by integrating this equation of the first order. If the two factors are distinct, it may happen that the two systems which arise may both be integrated. Then two first integrals $u_1 = \Phi(u_2)$ and $v_1 = \Psi(v_2)$ will be found, and instead of integrating one of these equations it may be better to solve both for p and q and to substitute in the expression $dz = pdx + qdy$ and integrate. When, however, it is not possible to find even one first integral, Monge's method fails.

As an example take $(x+y)(r-t)=-4p$. The equations are

$$(x+y)dy^2-(x+y)dx^2=0 \quad \text{or} \quad dy-dx=0, \qquad dy+dx=0$$

and
$$(x+y)dydp-(x+y)dxdq+4pdxdy=0. \tag{A}$$

Now the equation $dy-dx=0$ may be integrated at once to give $y=x+C_1$. The second equation (A) then takes the form

$$2xdp+4pdx-2xdq+C_1(dp-dq)=0;$$

but as $dz=pdx+qdy=(p+q)dx$ in this case, we have by combination

$$2(xdp+pdx)-2(xdq+qdx)+C_1(dp-dq)+2dz=0$$

or
$$(2x+C_1)(p-q)+2z=C_2 \quad \text{or} \quad (x+y)(p-q)+2z=C_2.$$

Hence
$$(x+y)(p-q)+2z=\Phi(y-x) \tag{27}$$

is a first integral. This is linear and may be integrated by

$$\frac{dx}{x+y}=-\frac{dy}{x+y}=\frac{dz}{\Phi(y-x)-2z} \quad \text{or} \quad x+y=K_1, \qquad \frac{dx}{K_1}=\frac{dz}{\Phi(K_1-2x)-2z}.$$

This equation is an ordinary linear equation in z and x. The integration gives

$$K_1ze^{\frac{2x}{K_1}}=\int e^{\frac{2x}{K_1}}\Phi(K_1-2x)dx+K_2.$$

Hence
$$(x+y)ze^{\frac{2x}{x+y}}-\int e^{\frac{2x}{K_1}}\Phi(K_1-2x)\,dx=K_2=\Psi(K_1)=\Psi(x+y)$$

is the general integral of the given equation when K_1 has been replaced by $x+y$ after integration, — an integration which cannot be performed until Φ is given.

The other method of solution would be to use also the second system containing $dy+dx=0$ instead of $dy-dx=0$. Thus in addition to the first integral (27) a second intermediary integral might be sought. The substitution of $dy+dx=0$, $y+x=C_1$ in (A) gives $C_1(dp+dq)+4pdx=0$. This equation is not integrable, because $dp+dq$ is a perfect differential and pdx is not. The combination with $dz=pdx+qdy=(p-q)dx$ does not improve matters. Hence it is impossible to determine a second intermediary integral, and the method of completing the solution by integrating (27) is the only available method.

Take the equation $ps-qr=0$. Here $S=p$, $R=-q$, $T=V=0$. Then

$$-qdy^2-pdxdy=0 \quad \text{or} \quad dy=0, \quad pdx+qdy=0 \quad \text{and} \quad -qdydp=0$$

are the equations to work with. The system $dy=0$, $qdydp=0$, $dz=pdx+qdy$, and the system $pdx+qdy=0$, $qdydp=0$, $dz=pdx+qdy$ are not very satisfactory for obtaining an intermediary integral $u_1=\Phi(u_2)$, although $p=\Phi(z)$ is an obvious solution of the first set. It is better to use a method adapted to this special equation. Note that

$$\frac{\partial}{\partial x}\left(\frac{q}{p}\right)=\frac{ps-qr}{p^2}, \quad \text{and} \quad \frac{\partial}{\partial x}\left(\frac{q}{p}\right)=0 \quad \text{gives} \quad \frac{q}{p}=f(y).$$

By (11), p. 124,
$$\frac{q}{p}=-\left(\frac{\partial x}{\partial y}\right)_z; \quad \text{then} \quad \frac{\partial x}{\partial y}=-f(y)$$

and
$$x=-\int f(y)\,dy+\Psi(z)=\Phi(y)+\Psi(z).$$

EXERCISES

1. Integrate these equations and discuss the singular solution:

(α) $p^{\frac{1}{2}} + q^{\frac{1}{2}} = 2x$, (β) $(p^2 + q^2)x = pz$, (γ) $(p+q)(px+qy) = 1$,
(δ) $pq = px + qy$, (ϵ) $p^2 + q^2 = x + y$, (ζ) $xp^2 - 2zp + xy = 0$,
(η) $q^2 = z^2(p-q)$, (θ) $q(p^2z + q^2) = 1$, (ι) $p(1+q^2) = q(z-c)$,
(κ) $xp(1+q) = qz$, (λ) $y^2(p^2-1) = x^2p^2$, (μ) $z^2(p^2+q^2+1) = c^2$,
(ν) $p = (z+yq)^2$, (o) $pz = 1 + q^2$, (π) $z - pq = 0$, (ρ) $q = xp + p^2$.

2. Show that the rule for the type of Ex. 13, p. 273, can be deduced by Charpit's method. How about the generalized Clairaut form of Ex. 15?

3. (α) For the solution of the type $f_1(x, p) = f_2(y, q)$, the rule is: Set

$$f_1(x, p) = f_2(y, q) = a,$$

and solve for p and q as $p = g_1(x, a)$, $q = g_2(y, a)$; the complete solution is

$$z = \int g_1(x, a)\,dx + \int g_2(y, a)\,dy + b.$$

(β) For the type $F(z, p, q) = 0$ the rule is: Set $X = x + ay$, solve

$$F\left(z, \frac{dz}{dX}, a\frac{dz}{dX}\right) \quad \text{for} \quad \frac{dz}{dX} = \phi(z, a), \quad \text{and let} \quad \int \frac{dz}{\phi(z, a)} = f(z, a);$$

the complete solution is $x + ay + b = f(z, a)$. Discuss these rules in the light of Charpit's method. Establish a rule for the type $F(x + y, p, q) = 0$. Is there any advantage in using the rules over the use of the general method? Assort the examples of Ex. 1 according to these rules as far as possible.

4. What is obtainable for partial differential equations out of any characteristics of homogeneity that may be present?

5. By differentiating $p = f(x, y, z, q)$ successively with respect to x and y show that the expansion of the solution by Taylor's Formula about the point (x_0, y_0, z_0) may be found if the successive derivatives with respect to y alone,

$$\frac{\partial z}{\partial y}, \quad \frac{\partial^2 z}{\partial y^2}, \quad \frac{\partial^3 z}{\partial y^3}, \quad \ldots, \quad \frac{\partial^n z}{\partial y^n}, \quad \ldots,$$

are assigned arbitrary values at that point. Note that this arbitrariness allows the solution to be passed through any curve through (x_0, y_0, z_0) in the plane $x = x_0$.

6. Show that $F(x, y, z, p, q) = 0$ satisfies Charpit's equations

$$du = \frac{dx}{-F'_p} = \frac{dy}{-F'_q} = \frac{dz}{-(pF'_p + qF'_q)} = \frac{dp}{F'_x + pF'_z} = \frac{dq}{F'_y + qF'_z}, \tag{28}$$

where u is an auxiliary variable introduced for symmetry. Show that the first three equations are the differential equations of the lineal elements of the cones of Ex. 6, p. 272. The integrals of (28) therefore define a system of curves which have a planar element of the equation $F = 0$ passing through each of their lineal tangential elements. If the equations be integrated and the results be solved for the variables, and if the constants be so determined as to specify one particular curve with the initial conditions x_0, y_0, z_0, p_0, q_0, then

$$x = x(u, x_0, y_0, z_0, p_0, q_0), \quad y = y(\cdots), \quad z = z(\cdots), \quad p = p(\cdots), \quad q = q(\cdots).$$

Note that, along the curve, $q = f(p)$ and that consequently the planar elements just mentioned must lie upon a developable surface containing the curve (§ 67). The curve and the planar elements along it are called a characteristic and a *characteristic strip* of the given differential equation. In the case of the linear equation the characteristic curves afforded the integration and any planar element through their lineal tangential elements satisfied the equation; but here it is only those planar elements which constitute the characteristic strip that satisfy the equation. What the complete integral does is to piece the characteristic strips into a family of surfaces dependent on two parameters.

7. By simple devices integrate the equations. Check the answers:

$(\alpha)\ \dfrac{\partial^2 z}{\partial x^2} = f(x)$, $\quad(\beta)\ \dfrac{\partial^n z}{\partial y^n} = 0$, $\quad(\gamma)\ \dfrac{\partial^2 z}{\partial x \partial y} = \dfrac{x}{y} + a$,

$(\delta)\ s + pf(x) = g(y)$, $\quad(\epsilon)\ ar = xy$, $\quad(\zeta)\ xr = (n-1)p$.

8. Integrate these equations by the method of factoring:

$(\alpha)\ (D_x^2 - a^2 D_y^2)z = 0$, $\quad(\beta)\ (D_x - D_y)^3 z = 0$, $\quad(\gamma)\ (D_x D_y^2 - D_y^3)z = 0$,

$(\delta)\ (D_x^2 + 3D_xD_y + 2D_y^2)z = x + y$, $\quad(\epsilon)\ (D_x^2 - D_xD_y - 6D_y^2)z = xy$,

$(\zeta)\ (D_x^2 - D_y^2 - 3D_x + 3D_y)z = 0$, $\quad(\eta)\ (D_x^2 - D_y^2 + 2D_x + 1)z = e^{-x}$.

9. Prove the operational equations:

$(\alpha)\ e^{\alpha x D_y}\phi(y) = (1 + \alpha x D_y + \frac{1}{2}\alpha^2 x^2 D_y^2 + \cdots)\phi(y) = \phi(y + \alpha x)$,

$(\beta)\ \dfrac{1}{D_x - \alpha D_y} 0 = e^{\alpha x D_y}\dfrac{1}{D_x} 0 = e^{\alpha x D_y}\phi(y) = \phi(y + \alpha x)$.

$(\gamma)\ \dfrac{1}{D_x - \alpha D_y} R(x, y) = e^{\alpha x D_y}\displaystyle\int^x e^{-\alpha\xi D_y} R(\xi, y)\,d\xi = \int^x R(\xi, y + \alpha x - \alpha\xi)\,d\xi$.

10. Prove that if $[(D_x - \alpha_1 D_y)^{m_1}\cdots(D_x - \alpha_k D_y)^{m_k}]z = 0$, then

$$z = \Phi_{11}(y + \alpha_1 x) + x\Phi_{12}(y + \alpha_1 x) + \cdots + x^{m_1 - 1}\Phi_{1m_1}(y + \alpha_1 x) + \cdots$$
$$+ \Phi_{k1}(y + \alpha_k x) + x\Phi_{k2}(y + \alpha_k x) + \cdots + x^{m_k - 1}\Phi_{km_k}(y + \alpha_k x).$$

where the Φ's are all arbitrary functions. This gives the solution of the reduced equation in the simplest case. What terms would correspond to $(D_x - \alpha D_y - \beta)^m z = 0$?

11. Write the solutions of the equations (or equations reduced) of Ex. 8.

12. State the rule of Ex. 9 (γ) as: Integrate $R(x, y - \alpha x)$ with respect to x and in the result change y to $y + \alpha x$. Apply this to obtaining particular solutions of Ex. 8 (δ), (ϵ), (η) with the aid of any short cuts that are analogous to those of Chap. VIII.

13. Integrate the following equations:

$(\alpha)\ (D_x^2 - D_{xy}^2 + D_y - 1)z = \cos(x + 2y) + e^y$, $\quad(\beta)\ x^2r^2 + 2xys + y^2t^2 = x^2 + y^2$,

$(\gamma)\ (D_x^2 + D_{xy} + D_y - 1)z = \sin(x + 2y)$, $\quad(\delta)\ r - t - 3p + 3q = e^{x+2y}$,

$(\epsilon)\ (D_x^3 - 2D_xD_y^2 + D_y^3)z = x^{-2}$, $\quad(\zeta)\ r - t + p + 3q - 2z = e^{x-y} - x^2y$.

$(\eta)\ (D_x^2 - D_xD_y - 2D_y^2 + 2D_x + 2D_y)z = e^{2x+3y} + \sin(2x + y) + xy$.

14. Try Monge's method on these equations of the second order:

$(\alpha)\ q^2r - 2pqs + p^2t = 0$, $\quad(\beta)\ r - a^2t = 0$, $\quad(\gamma)\ r + s = -p$.

$(\delta)\ q(1 + q)r - (p + q + 2pq)s + p(1 + p)t = 0$, $\quad(\epsilon)\ x^2r + 2xys + y^2t = 0$,

$(\zeta)\ (b + cq)^2r - 2(b + cq)(a + cp)s + (a + cp)^2t = 0$, $\quad(\eta)\ r + ka^2t = 2as$.

If any simpler method is available, state what it is and apply it also.

15. Show that an equation of the form $Rr + Ss + Tt + U(rt - s^2) = V$ necessarily arises from the elimination of the arbitrary function from

$$u_1(x, y, z, p, q) = f[u_2(x, y, z, p, q)].$$

Note that only such an equation can have an intermediary integral.

16. Treat the more general equation of Ex. 15 by the methods of the text and thus show that an intermediary integral may be sought by solving one of the systems

$$\begin{array}{ll} Udy + \lambda_1 Tdx + \lambda_1 Udp = 0, & Udx + \lambda_1 Rdy + \lambda_1 Udq = 0, \\ Udx + \lambda_2 Rdy + \lambda_2 Udq = 0, & Udy + \lambda_2 Tdx + \lambda_2 Udp = 0, \\ dz = pdx + qdy, & dz = pdx + qdy, \end{array}$$

where λ_1 and λ_2 are roots of the equation $\lambda^2(RT + UV) + \lambda US + U^2 = 0$.

17. Solve the equations: (α) $s^2 - rt = 0$, (β) $s^2 - rt = a^2$,
(γ) $ar + bs + ct + e(rt - s^2) = h$, (δ) $xqr + ypt + xy(s^2 - rt) = pq$.

PART III. INTEGRAL CALCULUS

CHAPTER XI

ON SIMPLE INTEGRALS

118. Integrals containing a parameter. Consider

$$\phi(\alpha) = \int_{x_0}^{x_1} f(x, \alpha)\, dx, \tag{1}$$

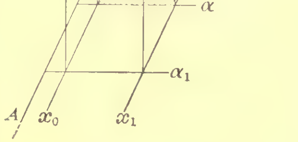
a definite integral which contains in the integrand a parameter α. If the indefinite integral is known, as in the case

$$\int \cos \alpha x dx = \frac{1}{\alpha} \sin \alpha x, \qquad \int_0^{\frac{\pi}{2}} \cos \alpha x dx = \frac{1}{\alpha} \sin \alpha x \Big|_0^{\frac{\pi}{2}} = \frac{1}{\alpha},$$

it is seen that the indefinite integral is a function of x and α, and that the definite integral is a function of α alone because the variable x disappears on the substitution of the limits. If the limits themselves depend on α, as in the case

$$\int_{\frac{1}{\alpha}}^{\alpha} \cos \alpha x dx = \frac{1}{\alpha} \sin \alpha x \Big|_{\frac{1}{\alpha}}^{\alpha} = \frac{1}{\alpha} (\sin \alpha^2 - \sin 1),$$

the integral is still a function of α.



In many instances the indefinite integral in (1) cannot be found explicitly and it then becomes necessary to discuss the continuity, differentiation, and integration of the function $\phi(\alpha)$ defined by the integral without having recourse to the actual evaluation of the integral; in fact these discussions may be required in order to effect that evaluation. Let the limits x_0 and x_1 be taken as constants independent of α. Consider the range of values $x_0 \leqq x \leqq x_1$ for x, and let $\alpha_0 \leqq \alpha \leqq \alpha_1$ be the range of values over which the function $\phi(\alpha)$ is to be discussed. The function $f(x, \alpha)$ may be plotted as the surface $z = f(x, \alpha)$ over the rectangle of values for (x, α). The

value $\phi(\alpha_i)$ of the function when $\alpha = \alpha_i$ is then the area of the section of this surface made by the plane $\alpha = \alpha_i$. If the surface $f(x, \alpha)$ is continuous, it is tolerably clear that the area $\phi(\alpha)$ will be continuous in α. *The function $\phi(\alpha)$ is continuous if $f(x, \alpha)$ is continuous in the two variables (x, α).*

To discuss the continuity of $\phi(\alpha)$ form the difference

$$\phi(\alpha + \Delta\alpha) - \phi(\alpha) = \int_{x_0}^{x_1} [f(x, \alpha + \Delta\alpha) - f(x, \alpha)]\, dx. \tag{2}$$

Now $\phi(\alpha)$ will be continuous if the difference $\phi(\alpha + \Delta\alpha) - \phi(\alpha)$ can be made as small as desired by taking $\Delta\alpha$ sufficiently small. If $f(x, y)$ is a continuous function of (x, y), it is possible to take Δx and Δy so small that the difference

$$|f(x + \Delta x, y + \Delta y) - f(x, y)| < \epsilon, \qquad |\Delta x| < \delta, \qquad |\Delta y| < \delta$$

for all points (x, y) of the region over which $f(x, y)$ is continuous (Ex. 3, p. 92). Hence in particular if $f(x, \alpha)$ be continuous in (x, α) over the rectangle, it is possible to take $\Delta\alpha$ so small that

$$|f(x, \alpha + \Delta\alpha) - f(x, \alpha)| < \epsilon, \qquad |\Delta\alpha| < \delta$$

for all values of x and α. Hence, by (65), p. 25,

$$|\phi(\alpha + \Delta\alpha) - \phi(\alpha)| = \left|\int_{x_0}^{x_1} [f(x, \alpha + \Delta\alpha) - f(x, \alpha)]\, dx\right| < \int_{x_0}^{x_1} \epsilon\, dx = \epsilon(x_1 - x_0).$$

It is therefore proved that the function $\phi(\alpha)$ is continuous provided $f(x, \alpha)$ is continuous in the two variables (x, α); for $\epsilon(x_1 - x_0)$ may be made as small as desired if ϵ may be made as small as desired.

As an illustration of a case where the condition for continuity is violated, take

$$\phi(\alpha) = \int_0^1 \frac{\alpha\, dx}{\alpha^2 + x^2} = \tan^{-1}\frac{x}{\alpha}\Big|_0^1 = \cot^{-1}\alpha \quad \text{if} \quad \alpha \neq 0, \quad \text{and} \quad \phi(0) = 0.$$

Here the integrand fails to be continuous for $(0, 0)$; it becomes infinite when $(x, \alpha) \doteq (0, 0)$ along any curve that is not tangent to $\alpha = 0$. The function $\phi(\alpha)$ is defined for all values of $\alpha \geqq 0$, is equal to $\cot^{-1}\alpha$ when $\alpha \neq 0$, and should therefore be equal to $\frac{1}{2}\pi$ when $\alpha = 0$ if it is to be continuous, whereas it is equal to 0. The importance of the imposition of the condition that $f(x, \alpha)$ be continuous is clear. It should not be inferred, however, that the function $\phi(\alpha)$ will *necessarily* be discontinuous when $f(x, \alpha)$ fails of continuity. For instance

$$\phi(\alpha) = \int_0^1 \frac{dx}{\sqrt{\alpha + x}} = \frac{1}{2}(\sqrt{\alpha + 1} - \sqrt{\alpha}), \qquad \phi(0) = \frac{1}{2}.$$

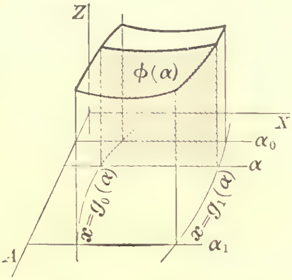
This function is continuous in α for all values $\alpha \geqq 0$; yet the integrand is discontinuous and indeed becomes infinite at $(0, 0)$. The condition of continuity imposed on $f(x, \alpha)$ in the theorem is *sufficient* to insure the continuity of $\phi(\alpha)$ but *by no means necessary*; when the condition is not satisfied some closer examination of the problem will sometimes disclose the fact that $\phi(\alpha)$ is still continuous.

In case the limits of the integral are functions of α, as

$$\phi(\alpha) = \int_{x_0 = g_0(\alpha)}^{x_1 = g_1(\alpha)} f(x, \alpha)\, dx, \qquad \alpha_0 \leqq \alpha \leqq \alpha_1, \tag{3}$$

the function $\phi(\alpha)$ will surely be continuous if $f(x, \alpha)$ is continuous over the region bounded by the lines $\alpha = \alpha_0$, $\alpha = \alpha_1$ and the curves $x_0 = g_0(\alpha)$, $x_1 = g_1(\alpha)$, and if the functions $g_0(\alpha)$ and $g_1(\alpha)$ are continuous.

For in this case



$$\phi(\alpha + \Delta\alpha) - \phi(\alpha) = \int_{g_0(\alpha + \Delta\alpha)}^{g_1(\alpha + \Delta\alpha)} f(x, \alpha + \Delta\alpha)\, dx$$

$$- \int_{g_0(\alpha)}^{g_1(\alpha)} f(x, \alpha)\, dx = \int_{g_0(\alpha + \Delta\alpha)}^{g_0(\alpha)} f(x, \alpha + \Delta\alpha)\, dx$$

$$+ \int_{g_1(\alpha)}^{g_1(\alpha + \Delta\alpha)} f(x, \alpha + \Delta\alpha)\, dx$$

$$+ \int_{g_0(\alpha)}^{g_1(\alpha)} [f(x, \alpha + \Delta\alpha) - f(x, \alpha)]\, dx.$$

The absolute values may be taken and the integrals reduced by (65), (65′), p. 25.

$$|\phi(\alpha + \Delta\alpha) - \phi(\alpha)| < \epsilon |g_1(\alpha) - g_0(\alpha)| + |f(\xi_1, \alpha + \Delta\alpha)|\,|\Delta g_1| + |f(\xi_0, \alpha + \Delta\alpha)|\,|\Delta g_0|,$$

where ξ_0 and ξ_1 are values of x between g_0 and $g_0 + \Delta g_0$, and g_1 and $g_1 + \Delta g_1$. By taking $\Delta\alpha$ small enough, $g_1(\alpha + \Delta\alpha) - g_1(\alpha)$ and $g_0(\alpha + \Delta\alpha) - g_0(\alpha)$ may be made as small as desired, and hence $\Delta\phi$ may be made as small as desired.

119. To find *the derivative of a function $\phi(\alpha)$ defined by an integral containing a parameter*, form the quotient

$$\frac{\Delta\phi}{\Delta\alpha} = \frac{\phi(\alpha + \Delta\alpha) - \phi(\alpha)}{\Delta\alpha}$$

$$= \frac{1}{\Delta\alpha}\left[\int_{g_0(\alpha + \Delta\alpha)}^{g_1(\alpha + \Delta\alpha)} f(x, \alpha + \Delta\alpha)\, dx - \int_{g_0(\alpha)}^{g_1(\alpha)} f(x, \alpha)\, dx\right],$$

$$\frac{\Delta\phi}{\Delta\alpha} = \int_{g_0(\alpha)}^{g_1(\alpha)} \frac{f(x, \alpha + \Delta\alpha) - f(x, \alpha)}{\Delta\alpha}\, dx + \int_{g_0 + \Delta g_0}^{g_0} \frac{f(x, \alpha + \Delta\alpha)}{\Delta\alpha}\, dx$$
$$+ \int_{g_1}^{g_1 + \Delta g_1} \frac{f(x, \alpha + \Delta\alpha)}{\Delta\alpha}\, dx.$$

The transformation is made by (53), p. 25. A further reduction may be made in the last two integrals by (65′), p. 25, which is the Theorem of the Mean for integrals, and the integrand of the first integral may be modified by the Theorem of the Mean for derivatives (p. 7, and Ex. 14, p. 10). Then

$$\frac{\Delta\phi}{\Delta\alpha} = \int_{g_0(\alpha)}^{g_1(\alpha)} f'_\alpha(x, \alpha + \theta\Delta\alpha)\, dx - f(\xi_0, \alpha + \Delta\alpha)\frac{\Delta g_0}{\Delta\alpha} + f(\xi_1, \alpha + \Delta\alpha)\frac{\Delta g_1}{\Delta\alpha}$$

and
$$\frac{d\phi}{d\alpha} = \int_{g_0(\alpha)}^{g_1(\alpha)} \frac{\partial f}{\partial \alpha}\, dx - f(g_0, \alpha)\frac{dg_0}{d\alpha} + f(g_1, \alpha)\frac{dg_1}{d\alpha}. \qquad (4)$$

A critical examination of this work shows that the derivative $\phi'(\alpha)$ exists and may be obtained by (4) in case f'_α exists and is continuous

in (x, α) and $g_0(\alpha)$, $g_1(\alpha)$ are differentiable. In the particular case that the limits g_0 and g_1 are constants, (4) reduces to *Leibniz's Rule*

$$\frac{d\phi}{d\alpha} = \frac{d}{d\alpha}\int_{x_0}^{x_1} f(x, \alpha)\,dx = \int_{x_0}^{x_1} \frac{\partial f}{\partial \alpha}\,dx, \tag{4'}$$

which states that *the derivative of a function defined by an integral with fixed limits may be obtained by differentiating under the sign of integration.* The additional two terms in (4), when the limits are variable, may be considered as arising from (66), p. 27, and Ex. 11, p. 30.

This process of *differentiating under the sign of integration is of frequent use in evaluating the function* $\phi(\alpha)$ in cases where the indefinite integral of $f(x, \alpha)$ cannot be found, but the indefinite integral of f'_α can be found. For if

$$\phi(\alpha) = \int_{x_0}^{x_1} f(x, \alpha)\,dx, \quad \text{then} \quad \frac{d\phi}{d\alpha} = \int_{x_1}^{x_1} f'_\alpha dx = \psi(\alpha).$$

Now an integration with respect to α will give ϕ as a function of α with a constant of integration which may be determined by the usual method of giving α some special value. Thus

$$\phi(\alpha) = \int_0^1 \frac{x^\alpha - 1}{\log x}\,dx, \qquad \frac{d\phi}{d\alpha} = \int_0^1 \frac{x^\alpha \log x}{\log x}\,dx = \int_0^1 x^\alpha dx.$$

Hence $$\frac{d\phi}{d\alpha} = \frac{1}{\alpha+1}x^{\alpha+1}\Big|_0^1 = \frac{1}{\alpha+1}, \qquad \phi(\alpha) = \log(\alpha+1) + C.$$

But $$\phi(0) = \int_0^1 0\,dx = 0 \quad \text{and} \quad \phi(0) = \log 1 + C.$$

Hence $$\phi(\alpha) = \int_0^1 \frac{x^\alpha - 1}{\log x}\,dx = \log(\alpha+1).$$

In the way of comment upon this evaluation it may be remarked that the functions $(x^\alpha - 1)/\log x$ and x^α are continuous functions of (x, α) for all values of x in the interval $0 \leqq x \leqq 1$ of integration and all positive values of α less than any assigned value, that is, $0 \leqq \alpha \leqq K$. The conditions which permit the differentiation under the sign of integration are therefore satisfied. This is not true for negative values of α. When $\alpha < 0$ the derivative x^α becomes infinite at (0, 0). The method of evaluation cannot therefore be applied without further examination. As a matter of fact $\phi(\alpha) = \log(\alpha+1)$ is defined for $\alpha > -1$, and it would be natural to think that some method could be found to justify the above formal evaluation of the integral when $-1 < \alpha \leqq K$ (see Chap. XIII).

To illustrate the application of the rule for differentiation when the limits are functions of α, let it be required to differentiate

$$\phi(\alpha) = \int_\alpha^{\alpha^2} \frac{x^\alpha - 1}{\log x}\,dx, \qquad \frac{d\phi}{d\alpha} = \int_\alpha^{\alpha^2} x^\alpha dx + \frac{\alpha^{2\alpha} - 1}{\log \alpha}\alpha - \frac{\alpha^\alpha - 1}{\log \alpha},$$

or $$\frac{d\phi}{d\alpha} = \frac{\alpha^{a+1}}{\alpha+1}\left[\alpha^{a+1} - 1\right] + \frac{1}{\log \alpha}\left[\alpha^{2\alpha} - \alpha^{\alpha} - \alpha + 1\right].$$

This formal result is only good subject to the conditions of continuity. Clearly α must be greater than zero. This, however, is the only restriction. It might seem at first as though the value $x = 1$ with $\log x = 0$ in the denominator of $(x^\alpha - 1)/\log x$ would cause difficulty; but when $x = 0$, this fraction is of the form $0/0$ and has a finite value which pieces on continuously with the neighboring values.

120. The next problem would be to find *the integral of a function defined by an integral containing a parameter.* The attention will be restricted to the case where the limits x_0 and x_1 are constants. Consider the integrals

$$\int_{a_0}^{a} \phi(a)\,da = \int_{\alpha_0}^{\alpha}\cdot\int_{x_0}^{x_1} f(x, \alpha)\,dx \cdot d\alpha,$$

where α may be any point of the interval $\alpha_0 \leqq \alpha \leqq \alpha_1$ of values over which $\phi(\alpha)$ is treated. Let

$$\Phi(\alpha) = \int_{x_0}^{x_1}\cdot\int_{\alpha_0}^{\alpha} f(x, \alpha)\,d\alpha \cdot dx.$$

Then $$\Phi'(\alpha) = \int_{x_0}^{x_1}\cdot\frac{\partial}{\partial\alpha}\int_{\alpha_0}^{\alpha} f(x, \alpha)\,d\alpha \cdot dx = \int_{x_0}^{x_1} f(x, \alpha)\,dx = \phi(\alpha)$$

by (4′), and by (66), p. 27; and the differentiation is legitimate if $f(x, a)$ be assumed continuous in (x, α). Now integrate with respect to α. Then

$$\int_{a_0}^{\alpha} \Phi'(\alpha) = \Phi(a) - \Phi(\alpha_0) = \int_{a_0}^{\alpha} \phi(\alpha)\,d\alpha.$$

But $\Phi(\alpha_0) = 0$. Hence, on substitution,

$$\Phi(a) = \int_{x_0}^{x_1}\cdot\int_{a_0}^{a} f(x, a)\,da \cdot dx = \int_{a_0}^{\alpha} \phi(\alpha)\,d\alpha = \int_{a_0}^{a}\cdot\int_{x_0}^{x_1} f(x, a)\,dx \cdot da. \quad (5)$$

Hence appears the rule for integration, namely, *integrate under the sign of integration.* The rule has here been obtained by a trick from the previous rule of differentiation; it could be proved directly by considering the integral as the limit of a sum.

It is interesting to note the interpretation of this integration on the figure, p. 281. As $\phi(\alpha)$ is the area of a section of the surface, the product $\phi(a)\,da$ is the infinitesimal volume under the surface and included between two neighboring planes. The integral of $\phi(\alpha)$ is therefore the volume * under the surface and boxed in by the four

* For the "volume of a solid with parallel bases and variable cross section" see Ex. 10, p. 10, and § 35 with Exs. 20, 23 thereunder.

planes $\alpha = \alpha_0$, $\alpha = \alpha_1$, $x = x_0$, $x = x_1$. The geometric significance of the reversal of the order of integrations, as

$$V = \int_{x_0}^{x_1} \cdot \int_{\alpha_0}^{\alpha_1} f(x, \alpha)\, d\alpha \cdot dx = \int_{\alpha_0}^{\alpha_1} \cdot \int_{x_0}^{x_1} f(x, \alpha)\, dx \cdot d\alpha,$$

is in this case merely that the volume may be regarded as generated by a cross section moving parallel to the $z\alpha$-plane, or by one moving parallel to the zx-plane, and that the evaluation of the volume may be made by either method. If the limits x_0 and x_1 depend on α, the integral of $\phi(\alpha)$ cannot be found by the simple rule of integration under the sign of integration. It should be remarked that integration under the sign may serve to evaluate functions defined by integrals.

As an illustration of integration under the sign in a case where the method leads to a function which may be considered as evaluated by the method, consider

$$\phi(\alpha) = \int_0^1 x^\alpha dx = \frac{1}{\alpha + 1}, \qquad \int_a^b \phi(\alpha)\, d\alpha = \int_a^b \frac{d\alpha}{\alpha + 1} = \log \frac{b+1}{a+1}.$$

But $\displaystyle\int_a^b \phi(\alpha)\, d\alpha = \int_0^1 \cdot \int_a^b x^\alpha d\alpha \cdot dx = \int_0^1 \frac{x^\alpha}{\log x}\Bigg|_{\alpha=a}^{\alpha=b} dx = \int_0^1 \frac{x^b - x^a}{\log x} dx.$

Hence $\displaystyle\int_0^1 \frac{x^b - x^a}{\log x} dx = \log \frac{b+1}{a+1} = \psi(a, b), \qquad a \geqq 0, \qquad b \geqq 0.$

In this case the integrand contains two parameters a, b, and the function defined is a function of the two. If $a = 0$, the function reduces to one previously found. It would be possible to repeat the integration. Thus

$$\int_0^1 \frac{x^\alpha - 1}{\log x} dx = \log(\alpha + 1). \qquad \int_0^\alpha \log(\alpha + 1)\, d\alpha = (\alpha + 1) \log(\alpha + 1) - \alpha.$$

$$\int_0^1 \cdot \int_0^\alpha \frac{x^\alpha - 1}{\log x} d\alpha \cdot dx = \int_0^1 \frac{x^\alpha - 1 - \alpha \log x}{(\log x)^2} dx = (\alpha + 1) \log(\alpha + 1) - \alpha$$

This is a new form. If here α be set equal to any number, say 1, then

$$\int_0^1 \frac{x - 1 - \log x}{(\log x)^2} dx = 2 \log 2 - 1.$$

In this way there has been evaluated a definite integral which depends on no parameter and which might have been difficult to evaluate directly. *The introduction of a parameter and its subsequent equation to a particular value is of frequent use in evaluating definite integrals.*

EXERCISES

1. Evaluate directly and discuss for continuity, $0 \leqq \alpha \leqq 1$:

$$(\alpha) \int_0^1 \frac{\alpha^2 dx}{\alpha^2 + x^2}, \qquad (\beta) \int_0^1 \frac{dx}{\sqrt{\alpha^2 + x^2}}, \qquad (\gamma) \int_0^1 \frac{x dx}{\sqrt{\alpha^2 + x^2}}.$$

2. If $f(x, \alpha, \beta)$ is a function containing two parameters and is continuous in the three variables (x, α, β) when $x_0 \leqq x \leqq x_1$, $\alpha_0 \leqq \alpha \leqq \alpha_1$, $\beta_0 \leqq \beta \leqq \beta_1$, show

$$\int_{x_0}^{x_1} f(x, \alpha, \beta)\, dx = \phi(\alpha, \beta) \text{ is continuous in } (\alpha, \beta).$$

3. Differentiate and hence evaluate and state the valid range for α:

$$(\alpha)\quad \int_0^\pi \log(1+\alpha\cos x)\,dx = \pi\log\frac{1+\sqrt{1-\alpha^2}}{2},$$

$$(\beta)\quad \int_0^\pi \log(1-2\alpha\cos x+\alpha^2)\,dx = \begin{cases}\pi\log\alpha^2, & \alpha^2 \geqq 1\\ 0, & \alpha^2 \leqq 1\end{cases}.$$

4. Find the derivatives without previously integrating:

$$(\alpha)\ \int_{\tan^{-1}\alpha}^{\sin^{-1}\alpha}\frac{1}{x}\tan\alpha x\,dx,\qquad (\beta)\ \int_0^{\alpha^2}\tan^{-1}\frac{x}{\alpha^2}\,dx,\qquad (\gamma)\ \int_{-ax}^{ax} e^{-\frac{b^2x^2}{a^2}}\,dx.$$

5. Extend the assumptions and the work of Ex. 2 to find the partial derivatives ϕ'_α and ϕ'_β and the total differential $d\phi$ if x_0 and x_1 are constants.

6. Prove the rule for integrating under the sign of integration by the direct method of treating the integral as the limit of a sum.

7. From Ex. 6 derive the rule for differentiating under the sign. Can the complete rule including the case of variable limits be obtained this way?

8. Note that the integral $\int_{x_0}^{g(x,\alpha)} f(x,\alpha)\,dx$ will be a function of (x,α). Derive formulas for the partial derivatives with respect to x and α.

9. Differentiate: $(\alpha)\ \dfrac{\partial}{\partial\alpha}\displaystyle\int_0^{ax}\sin(x+\alpha)\,dx,\quad (\beta)\ \dfrac{d}{dx}\displaystyle\int_0^{\sqrt[3]{x}} x^2dx.$

10. Integrate under the sign and hence evaluate by subsequent differentiation:

$$(\alpha)\ \int_0^1 x^a\log x\,dx,\qquad (\beta)\ \int_0^{\frac{\pi}{2}} x\sin\alpha x\,dx,\qquad (\gamma)\ \int_0^1 x\sec^2\alpha x\,dx.$$

11. Integrate or differentiate both sides of these equations:

$(\alpha)\ \displaystyle\int_0^1 x^a dx = \frac{1}{a+1}$ to show $\displaystyle\int_0^1 x^a(\log x)^n dx = (-1)^n\frac{n!}{(a+1)^{n+1}},$

$(\beta)\ \displaystyle\int_0^\infty \frac{dx}{x^2+\alpha} = \frac{\pi}{2\sqrt{\alpha}}$ to show $\displaystyle\int_0^\infty \frac{dx}{(x^2+\alpha)^{n+1}} = \frac{\pi}{2}\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n\cdot\alpha^{n+\frac{1}{2}}},$

$(\gamma)\ \displaystyle\int_0^\infty e^{-ax}\cos mx\,dx = \frac{\alpha}{\alpha^2+m^2}$ to show $\displaystyle\int_0^\infty \frac{e^{-\alpha x}-e^{-\beta x}}{x\sec mx}dx = \frac{1}{2}\log\left(\frac{\beta^2+m^2}{\alpha^2+m^2}\right),$

$(\delta)\ \displaystyle\int_0 e^{-ax}\sin mx\,dx = \frac{m}{\alpha^2+m^2}$ to show $\displaystyle\int_0^\infty \frac{e^{-\alpha x}-e^{-\beta x}}{x\csc mx}dx = \tan^{-1}\frac{\beta}{m}-\tan^{-1}\frac{\alpha}{m},$

$(\epsilon)\ \displaystyle\int_0^\pi \frac{dx}{a-\cos x} = \frac{\pi}{\sqrt{a^2-1}}$ to find $\displaystyle\int_0^\pi \frac{dx}{(a-\cos x)^2},\ \int_0^\pi \log\frac{b-\cos x}{a-\cos x},$

$(\zeta)\ \displaystyle\int_0^\infty \frac{x^{a-1}dx}{1+x} = \frac{\pi}{\sin\pi a}$ to find $\displaystyle\int_0^\infty \frac{x^{a-1}\log x\,dx}{1+x},\ \int_0^\infty \frac{x^{b-1}-x^{a-1}}{(1+x)\log x}dx.$

Note that in (β)–(δ) the integrals extend to infinity and that, as the rules of the text have been proved on the hypothesis that the interval of integration is finite, a further justification for applying the rules is necessary; this will be treated in Chap. XIII, but at this point the rules may be applied formally without justification.

12. Evaluate by any means these integrals:

$$(\alpha)\quad \int_0^a \sqrt{a^2 - x^2}\cos^{-1}\frac{x}{a}\,dx = a^2\left(\frac{\pi^2}{16} + \frac{1}{4}\right),$$

$$(\beta)\quad \int_0^{\frac{\pi}{2}} \frac{\log(1 + \cos\alpha\cos x)}{\cos x}\,dx = \frac{1}{2}\left(\frac{\pi^2}{4} - \alpha^2\right),$$

$$(\gamma)\quad \int_0^{\frac{\pi}{2}} \log(\alpha^2\cos^2 x + \beta^2\sin^2 x)\,dx = \pi\log\frac{\alpha + \beta}{2},$$

$$(\delta)\quad \int_0^{\infty} xe^{-\alpha x}\cos\beta x\,dx = \frac{\alpha^2 - \beta^2}{(\alpha^2 + \beta^2)^2},$$

$$(\epsilon)\quad \int_0^{\frac{\pi}{2}} \log\frac{a + b\sin x}{a - b\sin x}\,\frac{dx}{\sin x} = \pi\sin^{-1}\frac{b}{a}, \qquad b < a,$$

$$(\zeta)\quad \int_0^{\pi} \frac{\log(1 + k\cos x)}{\cos x}\,dx = \pi\sin^{-1}k.$$

$$(\theta)\quad \int_0^1 \log f(a + x)\,dx = \int_a^{a+1} \log f(x)\,dx = \int_0^a \log\frac{f(a+1)}{f(a)}\,da + \int_0^1 \log f(x)\,dx.$$

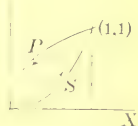
121. Curvilinear or line integrals. It is familiar that

$$A = \int_a^b y\,dx = \int_a^b f(x)\,dx$$

is the area between the curve $y = f(x)$, the x-axis, and the ordinates $x = a$, $x = b$. The formula may be used to evaluate more complicated areas. For instance, the area between the parabola $y^2 = x$ and the semicubical parabola $y^2 = x^3$ is

$$A = \int_0^1 x^{\frac{1}{2}}dx - \int_0^1 x^{\frac{3}{2}}dx = {}_P\!\int_0^1 y\,dx - {}_S\!\int_0^1 y\,dx,$$

where in the second expression the subscripts P and S denote that the integrals are evaluated for the parabola and semicubical parabola. As a change in the order of the limits changes the sign of the integral, the area may be written



$$A = {}_P\!\int_0^1 y\,dx + {}_S\!\int_1^0 y\,dx = -{}_P\!\int_1^0 y\,dx - {}_S\!\int_0^1 y\,dx,$$

and is the area bounded by the closed curve formed of the portions of the parabola and semicubical parabola from 0 to 1.

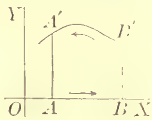
In considering the area bounded by a closed curve it is convenient to arrange the limits of the different integrals so that they follow the curve in a definite order. Thus if one advances along P from 0 to 1 and returns along S from 1 to 0, the entire closed curve has been described in a uniform direction and the inclosed area has been constantly on the right-hand side; whereas if one advanced along S from 0 to 1 and

returned from 1 to 0 along P, the curve would have been described in the opposite direction and the area would have been constantly on the left-hand side. Similar considerations apply to more general closed curves and lead to the definition: If a closed curve which nowhere crosses itself is described in such a direction as to keep the inclosed area always upon the left, the area is considered as positive; whereas if the description were such as to leave the area on the right, it would be taken as negative. It is clear that to a person standing in the inclosure and watching the description of the boundary, the description would appear counterclockwise or positive in the first case (§ 76).

In the case above, the area when positive is

$$A = -\left[\int_{S\,0}^{1} y\,dx + \int_{P\,1}^{0} y\,dx\right] = -\int_{\bigcirc} y\,dx. \tag{6}$$

where in the last integral the symbol $\bigcirc$ denotes that the integral is to be evaluated around the closed curve by describing the curve in the positive direction. That the formula holds for the ordinary case of area under a curve may be verified at once. Here the circuit consists of the contour $ABB'A'A$. Then



$$\int_{\bigcirc} y\,dx = \int_{A}^{B} y\,dx + \int_{B}^{B'} y\,dx + \int_{B'}^{A'} y\,dx + \int_{A'}^{A} y\,dx.$$

The first integral vanishes because $y = 0$, the second and fourth vanish because x is constant and $dx = 0$. Hence

$$-\int_{\bigcirc} y\,dx = -\int_{B'}^{A'} y\,dx = \int_{A'}^{B'} y\,dx.$$

It is readily seen that the two new formulas

$$A = \int_{\bigcirc} x\,dy \quad \text{and} \quad A = \tfrac{1}{2}\int_{\bigcirc} (x\,dy - y\,dx) \tag{7}$$

also give the area of the closed curve. The first is proved as (6) was proved and the second arises from the addition of the two. Any one of the three may be used to compute the area of the closed curve; the last has the advantage of symmetry and is particularly useful in finding the area of a sector, because along the lines issuing from the origin $y : x = dy : dx$ and $x\,dy - y\,dx = 0$; the previous form with the integrand $x\,dy$ is advantageous when part of the contour consists of lines parallel to the x-axis so that $dy = 0$; the first form has similar advantages when parts of the contour are parallel to the y-axis.

The connection of the third formula with the vector expression for the area is noteworthy. For (p. 175)

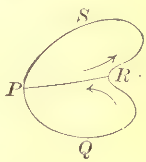
$$d\mathbf{A} = \tfrac{1}{2}\,\mathbf{r}\times d\mathbf{r}, \qquad \mathbf{A} = \tfrac{1}{2}\int_{\bigcirc} \mathbf{r}\times d\mathbf{r},$$

and if $$\mathbf{r} = x\mathbf{i} + y\mathbf{j}, \qquad d\mathbf{r} = \mathbf{i}\,dx + \mathbf{j}\,dy,$$

then $$\mathbf{A} = \int_{\bigcirc} \mathbf{r}\times d\mathbf{r} = \tfrac{1}{2}\,\mathbf{k}\int_{\bigcirc}(x\,dy - y\,dx).$$

The unit vector **k** merely calls attention to the fact that the area lies in the xy-plane perpendicular to the z-axis and is described so as to appear positive.

These formulas for the area as a curvilinear integral taken around the boundary have been derived from a simple figure whose contour was cut in only two points by a line parallel to the axes. The extension to more complicated contours is easy. In the first place note that if two closed areas are contiguous over a part of their contours, the integral around the total area following both contours, but omitting the part in common, is equal to the sum of the integrals. For



$$\int_{PRSP} + \int_{PQRP} = \int_{PR} + \int_{RSP} + \int_{PQR} + \int_{RP} = \int_{QRSP},$$

since the first and last integrals of the four are in opposite directions along the same line and must cancel. But the total area is also the sum of the individual areas and hence the integral around the contour $PQRSP$ must be the total area. The formulas for determining the area of a closed curve are therefore applicable to such areas as may be composed of a finite number of areas each bounded by an oval curve.

If the contour bounding an area be expressed in parametric form as $x = f(t)$, $y = \phi(t)$, the area may be evaluated as

$$\int f(t)\phi'(t)\,dt = -\int \phi(t)f'(t)\,dt = \tfrac{1}{2}\int [f(t)\phi'(t) - \phi(t)f'(t)]\,dt, \tag{7'}$$

where the limits for t are the value of t corresponding to any point of the contour and the value of t corresponding to the same point after the curve has been described once in the positive direction. Thus in the case of the strophoid

$$y^2 = x^2\frac{a - x}{a + x}, \quad \text{the line} \quad y = tx$$

cuts the curve in the double point at the origin and in only one other point; the coördinates of a point on the curve may be expressed as rational functions

$$x = a(1 - t^2)/(1 + t^2), \qquad y - at(1 - t^2)/(1 + t^2)$$

of t by solving the strophoid with the line: and when t varies from -1 to $+1$ the point (x, y) describes the loop of the strophoid and the limits for t are -1 and $+1$.

122. Consider next the meaning and the evaluation of

$$\int_{C\,a,b}^{x,y} [P(x, y)\,dx + Q(x, y)\,dy], \quad \text{where} \quad y = f(x). \tag{8}$$

This is called *a curvilinear or line integral along the curve* C or $y = f(x)$ from the point (a, b) to (x, y). It is possible to eliminate y by the relation $y = f(x)$ and write

$$\int_a^x [P(x, f(x)) + Q(x, f(x))\,f'(x)]\,dx. \tag{9}$$

The integral then becomes an ordinary integral in x alone. If the curve had been given in the form $x = f(y)$, it would have been better to convert the line integral into an integral in y alone. *The method of evaluating the integral is therefore defined.* The differential of the integral may be written as

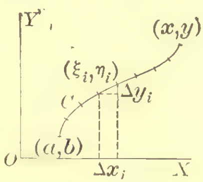
$$d\int_{a,b}^{x,y} (P\,dx + Q\,dy) = P\,dx + Q\,dy, \tag{10}$$

where either x and dx or y and dy may be eliminated by means of the equation of the curve C. For further particulars see § 123.

To get at *the meaning of the line integral*, it is necessary to consider it as the limit of a sum (compare § 16). Suppose that the curve C between (a, b) and (x, y) be divided into n parts, that Δx_i and Δy_i are the increments corresponding to the ith part, and that (ξ_i, η_i) is any point in that part. Form the sum

$$\sigma = \sum [P(\xi_i, \eta_i)\,\Delta x_i + Q(\xi_i, \eta_i)\,\Delta y_i]. \tag{11}$$

If, when n becomes infinite so that Δx and Δy each approaches 0 as a limit, the sum σ approaches a definite limit independent of how the individual increments Δx_i and Δy_i approach 0, and of how the point (ξ_i, η_i) is chosen in its segment of the curve, then this limit is defined as the line integral



$$\lim \sigma = \int_{C\,a,b}^{x,y} [P(x, y)\,dx + Q(x, y)\,dy]. \tag{12}$$

It should be noted that, as in the case of the line integral which gives the area, any line integral which is to be evaluated along two curves which have in common a portion described in opposite directions may be replaced by the integral along so much of the curves as not repeated; for the elements of σ corresponding to the common portion are equal and opposite.

That σ does approach a limit provided P and Q are continuous functions of (x, y) and provided the curve C is monotonic, that is, that neither Δx nor Δy changes its sign, is easy to prove. For the expression for σ may be written

$$\sigma = \sum \left[P(\xi_i, f(\xi_i))\,\Delta x_i + Q(f^{-1}(\eta_i), \eta_i)\,\Delta y_i\right]$$

by using the equation $y = f(x)$ or $x = f^{-1}(y)$ of C. Now as

$$\int_a^x P(x, f(x))\,dx \quad \text{and} \quad \int_b^y Q\,(f^{-1}(y), y)\,dy$$

are both existent ordinary definite integrals in view of the assumptions as to continuity, the sum σ must approach their sum as a limit. It may be noted that this proof does not require the continuity or existence of $f'(x)$ as does the formula (9). In practice the added generality is of little use. The restriction to a monotonic curve may be replaced by the assumption of a curve C which can be regarded as made up of a finite number of monotonic parts including perhaps some portions of lines parallel to the axes. More general varieties of C are admissible, but are not very useful in practice (§ 127).

Further to examine the line integral and appreciate its utility for mathematics and physics consider some examples. Let

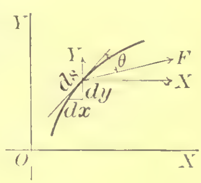
$$F(x, y) = X(x, y) + iY(x, y)$$

be a complex function (§ 73). Then

$$\begin{aligned} {}_C\!\int_{z=c}^{z=z} F(x, y)\,dz &= {}_C\!\int_{a,b}^{x,y} [X(x, y) + iY(x, y)][dx + idy] \\ &= {}_C\!\int_{a,b}^{x,y} (Xdx - Ydy) + i\,{}_C\!\int_{a,b}^{x,y} (Ydx + Xdy). \end{aligned} \tag{13}$$

It is apparent that *the integral of the complex function is the sum of two line integrals in the complex plane.* The value of the integral can be computed only by the assumption of some definite path C of integration and will differ for different paths (but see § 124).

By definition *the work done by a constant force F* acting on a particle, which moves a distance s along a straight line inclined at an angle θ to the force, is $W = Fs\cos\theta$. If the path were curvilinear and the force were variable, *the differential of work* would be taken as $dW = F\cos\theta ds$, where ds is the infinitesimal arc and θ is the angle between the arc and the force. Hence



$$W = \int dW = \int_{a,b}^{x,y} F\cos\theta ds = \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{F}\cdot d\mathbf{r},$$

where the path must be known to evaluate the integral and where the last expression is merely the equivalent of the others when the

notations of vectors are used (p. 164). These expressions may be converted into the ordinary form of the line integral. For

$$\mathbf{F} = X\mathbf{i} + Y\mathbf{j}, \qquad d\mathbf{r} = \mathbf{i}dx + \mathbf{j}dy, \qquad \mathbf{F}\cdot d\mathbf{r} = Xdx + Ydy,$$

and
$$W = \int_{a,b}^{x,y} F\cos\theta ds = \int_{a,b}^{x,y}(Xdx + Ydy),$$

where X and Y are the components of the force along the axes. It is readily seen that any line integral may be given this same interpretation. If

$$I = \int_{a,b}^{x,y} Pdx + Qdy, \quad \text{form} \quad \mathbf{F} = P\mathbf{i} + Q\mathbf{j}.$$

Then
$$I = \int_{a,b}^{x,y} Pdx + Qdy = \int_{a,b}^{x,y} F\cos\theta ds.$$

To the principles of momentum and moment of momentum (§ 80) may now be added the principle of work and energy for mechanics. Consider

$$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{F} \quad \text{and} \quad m\frac{d^2\mathbf{r}}{dt^2}\cdot d\mathbf{r} = \mathbf{F}\cdot d\mathbf{r} = dW.$$

Then
$$\frac{d}{dt}\left(\frac{1}{2}\frac{d\mathbf{r}}{dt}\cdot\frac{d\mathbf{r}}{dt}\right) = \frac{1}{2}\frac{d^2\mathbf{r}}{dt^2}\cdot\frac{d\mathbf{r}}{dt} + \frac{1}{2}\frac{d\mathbf{r}}{dt}\cdot\frac{d^2\mathbf{r}}{dt^2} = \frac{d^2\mathbf{r}}{dt^2}\cdot\frac{d\mathbf{r}}{dt},$$

or
$$d\left(\frac{1}{2}v^2\right) = \frac{d^2\mathbf{r}}{dt^2}\cdot d\mathbf{r} \quad \text{and} \quad d\left(\frac{1}{2}mv^2\right) = dW.$$

Hence
$$\frac{1}{2}mv^2 - \frac{1}{2}mv_0^2 = \int_{\mathbf{r}_0}^{\mathbf{r}}\mathbf{F}\cdot d\mathbf{r} = W.$$

In words: *The change of the kinetic energy $\frac{1}{2}mv^2$ of a particle moving under the action of the resultant force* $\mathbf{F}$ *is equal to the work done by the force*, that is, to the line integral of the force along the path. If there were several mutually interacting particles in motion, the results for the energy and work would merely be added as $\Sigma\frac{1}{2}mv^2 - \Sigma\frac{1}{2}mv_0^2 = \Sigma W$, and the total change in kinetic energy is the total work done by all the forces. The result gains its significance chiefly by the consideration of what forces may be disregarded in evaluating the work. As $dW = \mathbf{F}\cdot d\mathbf{r}$, the work done will be zero if $d\mathbf{r}$ is zero or if $\mathbf{F}$ and $d\mathbf{r}$ are perpendicular. Hence in evaluating W, forces whose point of application does not move may be omitted (for example, forces of support at pivots), and so may forces whose point of application moves normal to the force (for example, the normal reactions of smooth curves or surfaces). When more than one particle is concerned, the work done by the mutual actions and reactions may be evaluated as follows. Let $\mathbf{r}_1$, $\mathbf{r}_2$ be the vectors to the particles and $\mathbf{r}_1 - \mathbf{r}_2$ the vector joining them. The forces of action and reaction may be written as $\pm c(\mathbf{r}_1 - \mathbf{r}_2)$, as they are equal and opposite and in the line joining the particles. Hence

$$\begin{aligned} dW &= dW_1 + dW_2 = c(\mathbf{r}_1 - \mathbf{r}_2)\cdot d\mathbf{r}_1 - c(\mathbf{r}_1 - \mathbf{r}_2)\cdot d\mathbf{r}_2 \\ &= c(\mathbf{r}_1 - \mathbf{r}_2)\cdot d(\mathbf{r}_1 - \mathbf{r}_2) = \tfrac{1}{2}cd[(\mathbf{r}_1 - \mathbf{r}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2)] = \tfrac{1}{2}cdr_{12}^2, \end{aligned}$$

where r_{12} is the distance between the particles. Now dW vanishes when and only when dr_{12} vanishes, that is, when and only when the distance between the particles

remains constant. Hence *when a system of particles is in motion the change in the total kinetic energy in passing from one position to another is equal to the work done by the forces, where, in evaluating the work, forces acting at fixed points or normal to the line of motion of their points of application, and forces due to actions and reactions of particles rigidly connected, may be disregarded.*

Another important application is in the theory of thermodynamics. If U, p, v are the energy, pressure, volume of a gas inclosed in any receptacle, and if dU and dv are the increments of energy and volume when the amount dH of heat is added to the gas, then

$$dH = dU + pdv, \quad \text{and hence} \quad H = \int dU + pdv$$

is the total amount of heat added. By taking p and v as the independent variables,

$$H = \int \left[\frac{\partial U}{\partial p} dp + \left(\frac{\partial U}{\partial v} + p \right) dv \right] = \int [f(p, v)\,dp + g(p, v)\,dv].$$

The amount of heat absorbed by the system will therefore not depend merely on the initial and final values of (p, v) but on the sequence of these values between those two points, that is, upon the path of integration in the pv-plane.

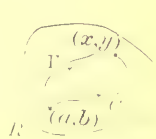
123. Let there be given a simply connected region (p. 89) bounded by a closed curve of the type allowed for line integrals, and let $P(x, y)$ and $Q(x, y)$ be continuous functions of (x, y) over this region. Then if the line integrals from (a, b) to (x, y) along two paths

$$\int_{C\,a,b}^{x,y} Pdx + Qdy = \int_{\Gamma\,a,b}^{x,y} Pdx + Qdy$$

are equal, the line integral taken around the combined path

$$\int_{C\,a,b}^{x,y} + \int_{\Gamma\,x,y}^{a,b} = \int Pdx + Qdy = 0$$

vanishes. This is a corollary of the fact that if the order of description of a curve is reversed, the signs of Δx_i and Δy_i and hence of the line integral are also reversed. Also, conversely, if the integral around the closed circuit is zero, the integrals from any point (a, b) of the circuit to any other point (x, y) are equal when evaluated along the two different parts of the circuit leading from (a, b) to (x, y).



The chief value of these observations arises in their application to the case where P and Q happen to be such functions that the line integral around any and every closed path lying in the region is zero. In this case if (a, b) be a fixed point and (x, y) be any point of the region, the line integral from (a, b) to (x, y) along any two paths lying within the region will be the same; for the two paths may be considered as forming one closed path, and the integral around that is zero by hypothesis. The value of the integral will therefore not depend at all on

the path of integration but only on the final point (x, y) to which the integration is extended. Hence the integral

$$\int_{a,b}^{x,y} [P(x, y)\,dx + Q(x, y)\,dy] = F(x, y), \tag{14}$$

extended from a fixed lower limit (a, b) to a variable upper limit (x, y), must be a function of (x, y).

This result may be stated as the theorem: *The necessary and sufficient condition that the line integral*

$$\int_{a,b}^{x,y} [P(x, y)\,dx + Q(x, y)\,dy]$$

define a single valued function of (x, y) over a simply connected region is that the circuit integral taken around any and every closed curve in the region shall be zero. This theorem, and in fact all the theorems on line integrals, may be immediately extended to the case of line integrals in space,

$$\int_{a,b,c}^{x,y,z} [P(x, y, z)\,dx + Q(x, y, z)\,dy + R(x, y, z)\,dz]. \tag{15}$$

If the integral about every closed path is zero so that the integral from a fixed lower limit to a variable upper limit

$$F(x, y) = \int_{a,b}^{x,y} P(x, y)\,dx + Q(x, y)\,dy$$

defines a function $F(x, y)$, that function has continuous first partial derivatives and hence a total differential, namely,

$$\frac{\partial F}{\partial x} = P, \qquad \frac{\partial F}{\partial y} = Q, \qquad dF = P\,dx + Q\,dy. \tag{16}$$

To prove this statement apply the definition of a derivative.

$$\frac{\partial F}{\partial x} = \lim_{\Delta x \doteq 0} \frac{\Delta F}{\Delta x} = \lim_{\Delta x \doteq 0} \frac{\displaystyle\int_{a,b}^{x+\Delta x,y} P\,dx + Q\,dy - \int_{a,b}^{x,y} P\,dx + Q\,dy}{\Delta x}.$$

Now as the integral is independent of the path, the integral to $(x + \Delta x, y)$ may follow the same path as that to (x, y), except for the passage from (x, y) to $(x + \Delta x, y)$ which may be taken along the straight line joining them. Then $\Delta y = 0$ and

$$\frac{\Delta F}{\Delta x} = \frac{1}{\Delta x}\int_{x,y}^{x+\Delta x,y} P(x, y)\,dx = \frac{1}{\Delta x} P(\xi, y)\,\Delta x = P(\xi, y),$$

by the Theorem of the Mean of (65′), p. 25. Now when $\Delta x \doteq 0$, the value ξ intermediate between x and $x + \Delta x$ will approach x and $P(\xi, y)$ will approach the limit $P(x, y)$ by virtue of its continuity. Hence $\Delta F/\Delta x$ approaches a limit and that limit is $P(x, y) = \partial F/\partial x$. The other derivative is treated in the same way.

If the integrand $Pdx + Qdy$ of a line integral is the total differential dF of a single valued function $F(x, y)$, then the integral about any closed circuit is zero and

$$\int_{a,b}^{x,y} Pdx + Qdy = \int_{a,b}^{x,y} dF = F(x, y) - F(a, b). \tag{17}$$

If equation (17) holds, it is clear that the integral around a closed path will be zero provided $F(x, y)$ is single valued; for $F(x, y)$ must come back to the value $F(a, b)$ when (x, y) returns to (a, b). If the function were not single valued, the conclusion might not hold.

To prove the relation (17), note that by definition

$$\int dF = \int Pdx + Qdy = \lim \sum [P(\xi_i, \eta_i)\Delta x_i + Q(\xi_i, \eta_i)\Delta y_i]$$

and $$\Delta F_i = P(\xi_i, \eta_i)\Delta x_i + Q(\xi_i, \eta_i)\Delta y_i + \epsilon_1 \Delta x_i + \epsilon_2 \Delta y_i,$$

where ϵ_1 and ϵ_2 are quantities which by the assumptions of continuity for P and Q may be made uniformly (§ 25) less than ϵ for all points of the curve provided Δx_i and Δy_i are taken small enough. Then

$$\left|\sum (P_i \Delta x_i + Q_i \Delta y_i) - \sum \Delta F_i\right| < \epsilon \sum (|\Delta x_i| + |\Delta y_i|);$$

and since $\Sigma \Delta F_i = F(x, y) - F(a, b)$, the sum $\Sigma P_i \Delta x_i + Q_i \Delta y_i$ approaches a limit, and that limit is

$$\lim \sum [P_i \Delta x_i + Q_i \Delta y_i] = \int_{a,b}^{x,y} Pdx + Qdy = F(x, y) - F(a, b).$$

EXERCISES

1. Find the area of the loop of the strophoid as indicated above.

2. Find, from (6), (7), the three expressions for the integrand of the line integrals which give the area of a closed curve in polar coördinates.

3. Given the equation of the ellipse $x = a \cos t$, $y = b \sin t$. Find the total area, the area of a segment from the end of the major axis to a line parallel to the minor axis and cutting the ellipse at a point whose parameter is t, also the area of a sector.

4. Find the area of a segment and of a sector for the hyperbola in its parametric form $x = a \cosh t$, $y = b \sinh t$.

5. Express the folium $x^3 + y^3 = 3axy$ in parametric form and find the area of the loop.

6. What area is given by the curvilinear integral around the perimeter of the closed curve $r = a \sin^3 \frac{1}{3}\phi$? What in the case of the lemniscate $r^2 = a^2 \cos 2\phi$ described as in making the figure 8 or the sign ∞?

7. Write for y the analogous form to (9) for x. Show that in curvilinear coördinates $x = \phi(u, v)$, $y = \psi(u, v)$ the area is

$$A = \frac{1}{2}\int \left[\begin{vmatrix} \phi & \psi \\ \phi'_u & \psi'_u \end{vmatrix} du + \begin{vmatrix} \phi & \psi \\ \phi'_v & \psi'_v \end{vmatrix} dv \right].$$

8. Compute these line integrals along the paths assigned:

(α) $\displaystyle\int_{0,\,0}^{1,\,1} x^2y\,dx + y^3dy$, $\quad y^2 = x \quad$ or $\quad y = x \quad$ or $\quad y^3 = x^2$,

(β) $\displaystyle\int_{0,\,0}^{1,\,1} (x^2 + y)\,dx + (x + y^2)\,dy$, $\quad y^2 = x \quad$ or $\quad y = x \quad$ or $\quad y^3 = x^2$,

(γ) $\displaystyle\int_{1,\,0}^{e,\,1} \frac{y}{x}\,dx + dy$, $\quad y = \log x \quad$ or $\quad y = 0 \quad$ and $\quad x = e$,

(δ) $\displaystyle\int_{0,\,0}^{x,\,y} x\sin y\,dx + y\cos x\,dy$, $\quad y = mx \quad$ or $\quad x = 0 \quad$ and $\quad y = y$,

(ϵ) $\displaystyle\int_{z=0}^{1-i} (x - iy)\,dz$, $\quad y = x \quad$ or $\quad x = 0 \quad$ and $\quad y = 1 \quad$ or $\quad y = 0 \quad$ and $\quad x = 1$.

(ζ) $\displaystyle\int_{z=1}^{z=i} (x^2 - (1 + i)\,xy + y^2)\,dz$, $\quad$ quadrant or straight line.

9. Show that $\int Pdx + Qdy = \int \sqrt{P^2 + Q^2}\cos\theta ds$ by working directly with the figure and without the use of vectors.

10. Show that if any circuit is divided into a number of circuits by drawing lines within it, as in a figure on p. 91, the line integral around the original circuit is equal to the sum of the integrals around the subcircuits taken in the proper order.

11. Explain the method of evaluating a line integral in space and evaluate:

(α) $\displaystyle\int_{0,\,0,\,0}^{1,\,1,\,1} x\,dx + 2\,y\,dy + z\,dz$, $\quad y^2 = x, \quad z^2 = x \quad$ or $\quad y = z = x$.

(β) $\displaystyle\int_{1,\,0,\,1}^{x,\,y,\,z} y\log x\,dx + y^2dy + \frac{x}{z}\,dz$, $\quad y = x - 1, \quad z = x^2 \quad$ or $\quad y = \log x, \quad z = x$.

12. Show that $\int Pdx + Qdy + Rdz = \int \sqrt{P^2 + Q^2 + R^2}\cos\theta ds$.

13. A bead of mass m strung on a frictionless wire of any shape falls from one point (x_0, y_0, z_0) to the point (x_1, y_1, z_1) on the wire under the influence of gravity. Show that $mg(z_0 - z_1)$ is the work done by all the forces, namely, gravity and the normal reaction of the wire.

14. If $x = f(t)$, $y = g(t)$, and $f'(t)$, $g'(t)$ be assumed continuous, show

$$\int_{a,\,b}^{x,\,y} P(x, y)\,dx + Q(x, y)\,dy = \int_{t_0}^{t}\left(P\frac{dx}{dt} + Q\frac{dy}{dt}\right)dt,$$

where $f(t_0) = a$ and $g(t_0) = b$. Note that this proves the statement made on page 290 in regard to the possibility of substituting in a line integral. The theorem is also needed for Exs. 1–8.

15. Extend to line integrals (15) in space the results of § 123.

16. *Angle as a line integral.* Show geometrically for a plane curve that $d\phi = \cos(r, n)\,ds/r$, where r is the radius vector of a curve and ds the element of

arc and (r, n) the angle between the radius produced and the normal to the curve, is the angle subtended at $r = 0$ by the element ds. Hence show that

$$\phi = \int \frac{\cos(r, n)}{r} ds = \int \frac{1}{r}\frac{dr}{dn} ds = \int \frac{d \log r}{dn} ds,$$

where the integrals are line integrals along the curve and dr/dn is the normal derivative of r, is the angle ϕ subtended by the curve at $r = 0$. Hence infer that

$$\int_{\circlearrowleft} \frac{d \log r}{dn} ds = 2\pi \quad \text{or} \quad \int_{\circlearrowleft} \frac{d \log r}{dn} ds = 0 \quad \text{or} \quad \int_{\circlearrowleft} \frac{d \log r}{dn} ds = \theta$$

according as the point $r = 0$ is within the curve or outside the curve or upon the curve at a point where the tangents in the two directions are inclined at the angle θ (usually π). Note that the formula may be applied at any point (ξ, η) if $r^2 = (\xi - x)^2 + (\eta - y)^2$ where (x, y) is a point of the curve. What would the integral give if applied to a space curve?

17. Are the line integrals of Ex. 16 of the same type $\int P(x, y)\,dx + Q(x, y)\,dy$ as those in the text, or are they more intimately associated with the curve? Cf. § 155.

18. Compute $(\alpha) \int_{1,0}^{0,1} (x - y)\,ds$, $(\beta) \int_{-1,0}^{0,1} xy\,ds$ along a right line, along a quadrant, along the axes.

124. Independency of the path. It has been seen that in case the integral around every closed path is zero or in case the integrand $Pdx + Qdy$ is a total differential, the integral is independent of the path, and conversely. Hence if

$$F(x, y) = \int_{a,b}^{x,y} Pdx + Qdy, \quad \text{then} \quad \frac{\partial F}{\partial x} = P, \qquad \frac{\partial F}{\partial y} = Q,$$

and
$$\frac{\partial^2 F}{\partial x \partial y} = \frac{\partial Q}{\partial x}, \qquad \frac{\partial^2 F}{\partial y \partial x} = \frac{\partial P}{\partial y}, \qquad \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x},$$

provided the partial derivatives P'_y and Q'_x are continuous functions.* It remains to prove the converse, namely, that: *If the two partial derivatives P'_y and Q'_x are continuous and equal, the integral*

$$\int_{a,b}^{x,y} Pdx + Qdy \quad \text{with} \quad P'_y = Q'_x \tag{18}$$

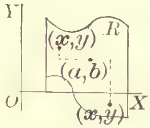
is independent of the path, is zero around a closed path, and the quantity $Pdx + Qdy$ is a total differential.

To show that the integral of $Pdx + Qdy$ around a closed path is zero if $P'_y = Q'_x$, consider first a region R such that any point (x, y) of it may

* See § 52. In particular observe the comments there made relative to differentials which are or which are not exact. This difference corresponds to integrals which are and which are not independent of the path.

be reached from (a, b) by following the lines $y = b$ and $x = x$. Then define the function $F(x, y)$ as

$$F(x, y) = \int_a^x P(x, b)\,dx + \int_b^y Q(x, y)\,dy \qquad (19)$$



for all points of that region R. Now

$$\frac{\partial F}{\partial y} = Q(x, y), \qquad \frac{\partial F}{\partial x} = P(x, b) + \frac{\partial}{\partial x}\int_b^y Q(x, y)\,dy.$$

But
$$\frac{\partial}{\partial x}\int_b^y Q(x, y)\,dy = \int_b^y \frac{\partial Q}{\partial x}\,dy = \int_b^y \frac{\partial P}{\partial y}\,dy = P(x, y)\Big]_b^y.$$

This results from Leibniz's rule (4′) of § 119, which may be applied since Q'_x is by hypothesis continuous, and from the assumption $Q'_x = P'_y$. Then

$$\frac{\partial F}{\partial x} = P(x, b) + P(x, y) - P(x, b) = P(x, y).$$

Hence it follows that, within the region specified, $Pdx + Qdy$ is the total differential of the function $F(x, y)$ defined by (19). Hence along any closed circuit within that region R the integral of $Pdx + Qdy$ is the integral of dF and vanishes.

It remains to remove the restriction on the type of region within which the integral around a closed path vanishes. Consider any closed path C which lies within the region over which P'_y and Q'_x are equal continuous functions of (x, y). As the path lies wholly within R it is possible to rule R so finely that any little rectangle which contains a portion of the path shall lie wholly within R. The reader may construct his own figure, possibly with reference to that of § 128, where a finer ruling would be needed. The path C may thus be surrounded by a zigzag line which lies within R. Each of the small rectangles within the zigzag line is a region of the type above considered and, by the proof above given, the integral around any closed curve within the small rectangle must be zero. Now the circuit C may be replaced by the totality of small circuits consisting either of the perimeters of small rectangles lying wholly within C or of portions of the curve C and portions of the perimeters of such rectangles as contain parts of C. And if C be so replaced, the integral around C is resolved into the sum of a large number of integrals about these small circuits; for the integrals along such parts of the small circuits as are portions of the perimeters of the rectangles occur in pairs with opposite signs.* Hence the integral around C is zero, where C is any circuit within R. Hence the integral of $Pdx + Qdy$ from (a, b) to (x, y) is independent of the path and defines a function $F(x, y)$ of which $Pdx + Qdy$ is the total differential. As this function is continuous, its value for points on the boundary of R may be defined as the limit of $F(x, y)$ as (x, y) approaches a point of the boundary, and it may thereby be seen that the line integral of (18) around the boundary is also 0 without any further restriction than that P'_y and Q'_x be equal and continuous within the boundary.

* See Ex. 10 above. It is well, in connection with §§ 123–125, to read carefully the work of §§ 44–45 dealing with varieties of regions, reducibility of circuits, etc.

It should be noticed that *the line integral*

$$\int_{a,b}^{x,y} Pdx + Qdy = \int_a^x P(x, b)\,dx + \int_b^y Q(x, y)\,dy, \tag{19}$$

when $Pdx + Qdy$ is an exact differential, that is, when $P'_y = Q'_x$, *may be evaluated by the rule given for integrating an exact differential* (p. 209), provided the path along $y = b$ and $x = x$ does not go outside the region. If that path should cut out of R, some other method of evaluation would be required. It should, however, be borne in mind that $Pdx + Qdy$ is best integrated by inspection whenever the function F, of which $Pdx + Qdy$ is the differential, can be recognized; if F is multiple valued, the consideration of the path may be required to pick out the particular value which is needed. It may be added that the work may be extended to line integrals in space without any material modifications.

It was seen (§ 73) that the conditions that the complex function

$$F(x, y) = X(x, y) + iY(x, y), \qquad z = x + iy,$$

be a function of the complex variable z are

$$X'_y = -Y'_x \text{ and } X'_x = Y'_y. \tag{20}$$

If these conditions be applied to the expression (13),

$$\int F(x, y) = \int_{a,b}^{x,y} Xdx - Ydy + i\int_{a,b}^{x,y} Ydx + Xdy,$$

for the line integral of such a function, it is seen that they are precisely the conditions (18) that each of the line integrals entering into the complex line integral shall be independent of the path. Hence *the integral of a function of a complex variable is independent of the path of integration in the complex plane, and the integral around a closed path vanishes.* This applies of course only to simply connected regions of the plane throughout which the derivatives in (20) are equal and continuous.

If the notations of vectors in three dimensions be adopted,

$$\int Xdx + Ydy + Zdz = \int \mathbf{F}\cdot d\mathbf{r},$$

where $\qquad \mathbf{F} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}, \qquad d\mathbf{r} = \mathbf{i}dx + \mathbf{j}dy + \mathbf{k}dz.$

In the particular case where the integrand is an exact differential and the integral around a closed path is zero,

$$Xdx + Ydy + Zdz = \mathbf{F}\cdot d\mathbf{r} = dU = d\mathbf{r}\cdot\nabla U,$$

where U is the function defined by the integral (for ∇U see p. 172). When $\mathbf{F}$ is interpreted as a force, the function $V = -U$ such that

$$\mathbf{F} = -\nabla V \quad \text{or} \quad X = -\frac{\partial V}{\partial x}, \qquad Y = -\frac{\partial V}{\partial y}, \qquad Z = -\frac{\partial V}{\partial z}$$

is called the potential function of the force $\mathbf{F}$. *The negative of the slope of the potential function is the force* $\mathbf{F}$ *and the negatives of the partial derivatives are the component forces along the axes.*

If the forces are such that they are thus derivable from a potential function, they are said to be *conservative*. In fact if

$$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{F} = -\nabla V, \qquad m\frac{d^2\mathbf{r}}{dt^2}\cdot d\mathbf{r} = -d\mathbf{r}\cdot\nabla V = -dV,$$

and
$$\int_{\mathbf{r}_0}^{\mathbf{r}_1} m\frac{d^2\mathbf{r}}{dt^2}\cdot d\mathbf{r} = \frac{m}{2}\frac{d\mathbf{r}}{dt}\cdot\frac{d\mathbf{r}}{dt}\Bigg|_{\mathbf{r}_0}^{\mathbf{r}_1} = -V\Bigg|_{\mathbf{r}_0}^{\mathbf{r}_1},$$

or
$$\frac{m}{2}(v_1^2 - v_0^2) = V_0 - V_1 \quad \text{or} \quad \frac{m}{2}v_1^2 + V_1 = \frac{m}{2}v_0^2 + V_0.$$

Thus the sum of the kinetic energy $\frac{1}{2}mv^2$ and the potential energy V is the same at all times or positions. This is the principle of the *conservation of energy* for the simple case of the motion of a particle when the force is conservative. In case the force is not conservative the integration may still be performed as

$$\frac{m}{2}(v_1^2 - v_0^2) = \int_{\mathbf{r}_0}^{\mathbf{r}_1} \mathbf{F}\cdot d\mathbf{r} = W,$$

where W stands for the work done by the force $\mathbf{F}$ during the motion. The result is that the change in kinetic energy is equal to the work done by the force; but dW is then not an exact differential and the work must not be regarded as a function of (x, y, z), — it depends on the path. The generalization to any number of particles as in § 123 is immediate.

125. The conditions that P'_y and Q'_x be continuous and equal, which insures independence of the path for the line integral of $Pdx + Qdy$, need to be examined more closely. Consider two examples:

First
$$\int Pdx + Qdy = \int \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy,$$

where
$$\frac{\partial P}{\partial y} = \frac{y^2-x^2}{(x^2+y^2)^2}, \qquad \frac{\partial Q}{\partial x} = \frac{y^2-x^2}{(x^2+y^2)^2}.$$

It appears formally that $P'_y = Q'_x$. If the integral be calculated around a square of side $2a$ surrounding the origin, the result is

$$\int_{-a}^{+a}\frac{+a\,dx}{x^2+a^2} + \int_{-a}^{+a}\frac{a\,dy}{a^2+y^2} + \int_{+a}^{-a}\frac{-a\,dx}{x^2+a^2} + \int_{+a}^{-a}\frac{-a\,dy}{a^2+y^2} = 2\int_{-a}^{+a}\frac{a\,dx}{x^2+a^2}$$
$$+ 2\int_{-a}^{+a}\frac{a\,dy}{a^2+y^2} = 4\int_{-a}^{+a}\frac{a\,d\xi}{\xi^2+a^2} = 4\frac{\pi}{2} = 2\pi \neq 0.$$

The integral fails to vanish around the closed path. The reason is not far to seek, the derivatives P'_y and Q'_x are not defined for (0, 0), and cannot be so defined as to be continuous functions of (x, y) near the origin. As a matter of fact

$$\int_{a,b}^{x,y} \frac{-ydx}{x^2+y^2} + \frac{xdy}{x^2+y^2} = \int_{a,b}^{x,y} d\tan^{-1}\frac{y}{x} = \tan^{-1}\frac{y}{x}\Big|_{a,b}^{x,y},$$

and $\tan^{-1}(y/x)$ is not a single valued function ; it takes on the increment 2π when one traces a path surrounding the origin (§ 45).

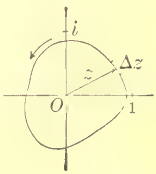
Another illustration may be found in the integral

$$\int \frac{dz}{z} = \int \frac{dx + idy}{x + iy} = \int \frac{xdx + ydy}{x^2+y^2} + i\int \frac{-ydx + xdy}{x^2+y^2}$$

taken along a path in the complex plane. At the origin $z = 0$ the integrand $1/z$ becomes infinite and so do the partial derivatives of its real and imaginary parts. If the integral be evaluated around a path passing once about the origin, the result is

$$\int_{\bigcirc} \frac{dz}{z} = \left[\frac{1}{2}\log(x^2+y^2) + i\tan^{-1}\frac{y}{x}\right]_{a,b}^{x,y} = 2\pi i. \tag{21}$$

In this case, as in the previous, the integral would necessarily be zero about any closed path which did not include the origin ; for then the conditions for absolute independence of the path would be satisfied. Moreover the integrals around two different paths each encircling the origin once would be equal ; for the paths may be considered as one single closed circuit by joining them with a line as in the device (§ 44) for making a multiply connected region simply connected, the integral around the complete circuit is zero, the parts due to the description of the line in the two directions cancel, and the integrals around the two given circuits taken in opposite directions are therefore equal and opposite. (Compare this work with the multiple valued nature of log z, p. 161.)



Suppose in general that $P(x, y)$ and $Q(x, y)$ are single valued functions which have the first partial derivatives P'_y and Q'_x continuous and equal over a region R except at certain points $A, B, \cdots$. Surround these points with small circuits. The remaining portion of R is such that P'_y and Q'_x are everywhere equal and continuous ; but the region is not simply connected, that is, it is possible to draw in the region circuits which cannot be shrunk down to a point, owing to the fact that the circuit may surround one or more of the regions which have been cut out. If a circuit can be shrunk down to a point, that is, if it is not inextricably wound about one or more of the deleted portions, the integral around the circuit will vanish ; for the previous reasoning will apply. But if the circuit coils about one or more of the deleted regions so that the attempt to shrink it down leads to a circuit which consists of the contours of these regions and of lines joining them, the integral need not vanish ; it reduces to the sum of a number of integrals

taken around the contours of the deleted portions. If one circuit can be shrunk into another, the integrals around the two circuits are equal if the direction of description is the same; for a line connecting the two circuits will give a combined circuit which can be shrunk down to a point.

The inference from these various observations is that in a multiply connected region the integral around a circuit need not be zero and the integral from a fixed lower limit (a, b) to a variable upper limit (x, y) may not be absolutely independent of the path, but may be different along two paths which are so situated relatively to the excluded regions that the circuit formed of the two paths from (a, b) to (x, y) cannot be shrunk down to a point. Hence

$$F(x, y) = \int_{a,b}^{x,y} Pdx + Qdy, \qquad P'_y = Q'_x \text{ (generally)},$$

the function defined by the integral, is not necessarily single valued. Nevertheless, any two values of $F(x, y)$ for the same end point will differ only by a sum of the form

$$F_2(x, y) - F_1(x, y) = m_1 I_1 + m_2 I_2 + \cdots$$

where $I_1, I_2, \ldots$ are the values of the integral taken around the contours of the excluded regions and where $m_1, m_2, \ldots$ are positive or negative integers which represent the number of times the combined circuit formed from the two paths will coil around the deleted regions in one direction or the other.

126. Suppose that $f(z) = X(x, y) + iY(x, y)$ is a single valued function of z over a region R surrounding the origin (see figure above), and that over this region the derivative $f'(z)$ is continuous, that is, the relations $X'_y = -Y'_x$ and $X'_x = Y'_y$ are fulfilled at every point so that no points of R need be cut out. Consider the integral

$$\int_O \frac{f(z)}{z} dz = \int_O \frac{X + iY}{x + iy} (dx + idy) \tag{22}$$

over paths lying within R. The function $f(z)/z$ will have a continuous derivative at all points of R except at the origin $z = 0$, where the denominator vanishes. If then a small circuit, say a circle, be drawn about the origin, the function $f(z)/z$ will satisfy the requisite conditions over the region which remains, and the integral (22) taken around a circuit which does not contain the origin will vanish.

The integral (22) taken around a circuit which coils once and only once about the origin will be equal to the integral taken around the

small circle about the origin. Now for the circle,

$$\int_{\odot}\frac{f(z)}{z}dz=\int_{\odot}\frac{f(0)+\eta(z)}{z}dz=f(0)\int_{\odot}\frac{dz}{z}+\int_{\odot}\frac{\eta}{z}dz,$$

where the assumed continuity of $f(z)$ makes $|\eta(z)|<\epsilon$ provided the circle about the origin is taken sufficiently small. Hence by (21)

$$\int_{\bigcirc}\frac{f(z)}{z}dz=\int_{\odot}\frac{f(z)}{z}dz=2\pi i f(0)+\xi$$

with $$|\xi|=\left|\int_{\odot}\frac{\eta}{z}dz\right|\leqq\int_{\odot}\left|\frac{\eta}{z}\right||dz|\leqq\epsilon\int_0^{2\pi}d\theta=2\pi\epsilon.$$

Hence the difference between (22) and $2\pi i f(0)$ can be made as small as desired, and as (22) is a certain constant, the result is

$$\int_{\bigcirc}\frac{f(z)}{z}dz=2\pi i f(0). \tag{23}$$

A function $f(z)$ which has a continuous derivative $f'(z)$ at every point of a region is said to be *analytic* over that region. Hence if the region includes the origin, the value of the analytic function at the origin is given by the formula

$$f(0)=\frac{1}{2\pi i}\int_{\bigcirc}\frac{f(z)}{z}dz, \tag{23'}$$

where the integral is extended over any circuit lying in the region and passing just once about the origin. It follows likewise that if $z=\alpha$ is any point within the region, then

$$f(\alpha)=\frac{1}{2\pi i}\int_{\bigcirc}\frac{f(z)}{z-\alpha}dz, \tag{24}$$

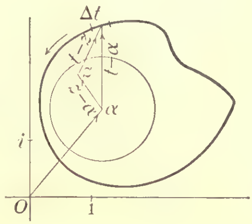
where the circuit extends once around the point α and lies wholly within the region. This important result is due to Cauchy.

A more convenient form of (24) is obtained by letting $t=z$ represent the value of z along the circuit of integration and then writing $\alpha=z$ and regarding z as variable. Hence Cauchy's Integral:

$$f(z)=\frac{1}{2\pi i}\int_{\bigcirc}\frac{f(t)}{t-z}dt. \tag{25}$$

This states that *if any circuit be drawn in the region over which $f(z)$ is analytic, the value of $f(z)$ at all points within that circuit may be obtained by evaluating Cauchy's Integral* (25). Thus $f(z)$ may be regarded

as defined by an integral containing a parameter z; for many purposes this is convenient. It may be remarked that when the values of $f(z)$ are given along any circuit, the integral may be regarded as defining $f(z)$ for all points within that circuit.



To find *the successive derivatives of* $f(z)$, it is merely necessary to differentiate with respect to z under the sign of integration. The conditions of continuity which are required to justify the differentiation are satisfied for all points z actually within the circuit and not upon it. Then

$$f''(z) = \frac{1}{2\pi i}\int_{\bigcirc}\frac{f(t)}{(t-z)^2}dt, \cdots, \; f^{(n-1)}(z) = \frac{(n-1)!}{2\pi i}\int_{\bigcirc}\frac{f(t)}{(t-z)^n}dt.$$

As the differentiations may be performed, these formulas show that *an analytic function has continuous derivatives of all orders.* The definition of the function only required a continuous first derivative.

Let α be any particular value of z (see figure). Then

$$\frac{1}{t-z} = \frac{1}{(t-\alpha)-(z-\alpha)} = \frac{1}{t-\alpha}\frac{1}{1-\dfrac{z-\alpha}{t-\alpha}}$$

$$= \frac{1}{t-\alpha}\left[1 + \frac{z-\alpha}{t-\alpha} + \frac{(z-\alpha)^2}{(t-\alpha)^2} + \cdots + \frac{(z-\alpha)^{n-1}}{(t-\alpha)^{n-1}} + \frac{\dfrac{(z-\alpha)^n}{(t-\alpha)^n}}{1-\dfrac{z-\alpha}{t-\alpha}}\right].$$

$$f(z) = \frac{1}{2\pi i}\int_{\bigcirc}\frac{f(t)}{t-z}dt = \frac{1}{2\pi i}\int_{\bigcirc}\frac{f(t)}{t-\alpha}dt + \frac{1}{2\pi i}\int_{\bigcirc}(z-\alpha)\frac{f(t)}{(t-\alpha)^2}dt$$

$$+ \frac{1}{2\pi i}\int_{\bigcirc}(z-\alpha)^2\frac{f(t)}{(t-\alpha)^3}dt + \cdots + \frac{1}{2\pi i}\int_{\bigcirc}(z-\alpha)^{n-1}\frac{f(t)}{(t-\alpha)^n}dt + R_n,$$

with
$$R_n = \frac{1}{2\pi i}\int_{\bigcirc}\frac{(z-\alpha)^n}{(t-\alpha)^n}\frac{1}{1-\dfrac{z-\alpha}{t-\alpha}}\frac{f(t)}{t-\alpha}dt.$$

Now t is the variable of integration and $z-\alpha$ is a constant with respect to the integration. Hence

$$f(z) = f(\alpha) + (z-\alpha)f'(\alpha) + \frac{(z-\alpha)^2}{2!}f''(\alpha) + \cdots + \frac{(z-\alpha)^{n-1}}{(n-1)!}f^{(n-1)}(\alpha) + R_n. \tag{26}$$

This is Taylor's Formula for a function of a complex variable.

EXERCISES

1. If $P'_y = Q'_x$, $Q'_z = R'_y$, $R'_x = P'_z$ and if these derivatives are continuous, show that $Pdx + Qdy + Rdz$ is a total differential.

2. Show that $\int_{C\,a,\,b}^{x,\,y} P(x,\,y,\,\alpha)\,dx + Q(x,\,y,\,\alpha)\,dy$, where C is a given curve, defines a continuous function of α, the derivative of which may be found by differentiating under the sign. What assumptions as to the continuity of P, Q, P'_α, Q'_α do you make?

3. If $\log z = \int_1^z \frac{dz}{z} = \int_{1,\,0}^{x,\,y} \frac{xdx + ydy}{x^2 + y^2} + i\int_{1,\,0}^{x,\,y} \frac{-ydx + xdy}{x^2 + y^2}$ be taken as the definition of $\log z$, draw paths which make $\log\left(\frac{1}{2} + \frac{1}{2}\sqrt{-3}\right) = \frac{1}{3}\pi i,\ 2\frac{1}{3}\pi i,\ -1\frac{2}{3}\pi i$.

4. Study $\int_0^z \frac{3z - 1}{z^2 - 1}$ with especial reference to closed paths which surround $+1$, -1, or both. Draw a closed path surrounding both and making the integral vanish.

5. If $f(z)$ is analytic for all values of z and if $|f(z)| < K$, show that

$$f(z) - f(0) = \int_{\bigcirc} f(t)\left[\frac{1}{t - z} - \frac{1}{t}\right]dt = \int_{\bigcirc} \frac{zf(t)}{(t - z)\,t}\,dt,$$

taken over a circle of large radius, can be made as small as desired. Hence infer that $f(z)$ must be the constant $f(z) = f(0)$.

6. If $G(z) = a_0 + a_1 z + \cdots + a_n z^n$ is a polynomial, show that $f(z) = 1/G(z)$ must be analytic over any region which does not include a root of $G(z) = 0$ either within or on its boundary. Show that the assumption that $G(z) = 0$ has no roots at all leads to the conclusion that $f(z)$ is constant and equal to zero. Hence infer that an algebraic equation has a root.

7. Show that the absolute value of the remainder in Taylor's Formula is

$$|R_n| = \frac{|z - a|^n}{2\pi}\left|\int_{\bigcirc} \frac{f(t)\,dt}{(t - a)^n\,(t - z)}\right| \leqq \frac{1}{2\pi}\,\frac{r^n}{\rho^n}\,\frac{ML}{\rho - r}$$

for all points z within a circle of radius r about a as center, when ρ is the radius of the largest circle concentric with a which can be drawn within the circuit about which the integral is taken, M is the maximum value of $f(t)$ upon the circuit, and L is the length of the circuit (figure above).

8. Examine for independence of path and in case of independence integrate:

$$(\alpha)\ \int x^2 y dx + xy^2 dy, \qquad (\beta)\ \int xy^2 dx + x^2 y dy, \qquad (\gamma)\ \int x dy + y dx,$$

$$(\delta)\ \int (x^2 + xy)\,dx + (y^2 + xy)\,dy, \qquad (\epsilon)\ \int y \cos x dy + \tfrac{1}{2}\,y^2 \sin x dx.$$

9. Find the conservative forces and the potential:

$$(\alpha)\ X = \frac{x}{(x^2 + y^2)^{\frac{3}{2}}},\quad Y = \frac{y}{(x^2 + y^2)^{\frac{3}{2}}},\quad Z = \frac{z}{(x^2 + y^2)^{\frac{3}{2}}},$$

$$(\beta)\ X = -nx,\ Y = -ny. \qquad (\gamma)\ X = 1/x,\ Y = y/x.$$

10. If $R(r, \phi)$ and $\Phi(r, \phi)$ are the component forces resolved along the radius vector and perpendicular to the radius, show that $dW = Rdr + r\Phi d\phi$ is the differential of work, and express the condition that the forces R, Φ be conservative.

11. Show that if a particle is acted on by a force $R = -f(r)$ directed toward the origin and a function of the distance from the origin, the force is conservative.

12. If a force follows the Law of Nature, that is, acts toward a point and varies inversely as the square r^2 of the distance from the point, show that the potential is $-k/r$.

13. From the results $\mathbf{F} = -\nabla V$ or $V = -\int \mathbf{F}\cdot d\mathbf{r} = \int Xdx + Ydy + Zdz$ show that if V_1 is the potential of $\mathbf{F}_1$ and V_2 of $\mathbf{F}_2$ then $V = V_1 + V_2$ will be the potential of $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$, that is, show that for conservative forces the addition of potentials is equivalent to the parallelogram law for adding forces.

14. If a particle is acted on by a retarding force $-k\mathbf{v}$ proportional to the velocity, show that $R = \frac{1}{2}kv^2$ is a function such that

$$\frac{\partial R}{\partial v_x} = -kv_x, \qquad \frac{\partial R}{\partial v_y} = -kv_y, \qquad \frac{\partial R}{\partial v_z} = -kv_z,$$
$$dW = -k\mathbf{v}\cdot d\mathbf{r} = -k(v_xdx + v_ydy + v_zdz).$$

Here R is called the dissipative function; show the force is not conservative.

15. Pick out the integrals independent of the path and integrate:

$$(\alpha)\ \int yzdx + xzdy + xydz, \qquad (\beta)\ \int ydx/z + xdy/z - xydz/z^2,$$

$$(\gamma)\ \int xyz(dx + dy + dz), \qquad (\delta)\ \int \log(xy)\,dx + xdy + ydz.$$

16. Obtain logarithmic forms for the inverse trigonometric functions, analogous to those for the inverse hyperbolic functions, either algebraically or by considering the inverse trigonometric functions as defined by integrals as

$$\tan^{-1}z = \int_0^z \frac{dz}{1+z^2}, \qquad \sin^{-1}z = \int_0^z \frac{dz}{\sqrt{1-z^2}}, \ldots.$$

17. Integrate these functions of the complex variable directly according to the rules of integration for reals and determine the values of the integrals by substitution:

$$(\alpha)\ \int_i^{1+i} ze^{2z^2}dz, \qquad (\beta)\ \int_0^{2i} \cos 3z\,dz, \qquad (\gamma)\ \int_1^{-1+i} (1+z^2)^{-1}dz,$$

$$(\delta)\ \int_0^{1+i} \frac{dz}{\sqrt{1-z^2}}, \qquad (\epsilon)\ \int_i^2 \frac{dz}{z\sqrt{z^2-1}}, \qquad (\zeta)\ \int_{-1}^{-2+i} \frac{dz}{\sqrt{1+z^2}}.$$

In the case of multiple valued functions mark two different paths and give two values.

18. Can the algorism of integration by parts be applied to the definite (or indefinite) integral of a function of a complex variable, it being understood that the integral must be a line integral in the complex plane? Consider the proof of Taylor's Formula by integration by parts, p. 57, to ascertain whether the proof is valid for the complex plane and what the remainder means.

19. Suppose that in a plane at $r = 0$ there is a particle of mass m which *attracts* according to the law $F = m/r$. Show that the potential is $V = m \log r$, so that $\mathbf{F} = -\nabla V$. The *induction* or *flux* of the force $\mathbf{F}$ outward across the element ds of a curve in the plane is by definition $-F\cos(F, n)\,ds$. By reference to Ex. 16, p. 297, show that the total induction or flux of $\mathbf{F}$ across the curve is the line integral (along the curve)

$$-\int F\cos(F, n)\,ds = m\int \frac{d\log r}{dn}\,ds = \int \frac{dV}{dn}\,ds;$$

and
$$m = \frac{-1}{2\pi}\int_{\bigcirc} F\cos(F, n)\,ds = \frac{1}{2\pi}\int_{\bigcirc} \frac{dV}{dn}\,ds,$$

where the circuit extends around the point $r = 0$, is a formula for obtaining the mass m within the circuit from the field of force $\mathbf{F}$ which is set up by the mass.

20. Suppose a number of masses $m_1, m_2, \cdots$, attracting as in Ex. 19, are situated at points $(\xi_1, \eta_1), (\xi_2, \eta_2), \cdots$ in the plane. Let

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \cdots, \qquad V = V_1 + V_2 + \cdots, \qquad V_i = m_i \log\left[(\xi_i - x)^2 + (\eta_i - y)^2\right]^{\frac{1}{2}}$$

be the force and potential at (x, y) due to the masses. Show that

$$\frac{-1}{2\pi}\int_{\bigcirc} F\cos(F, n)\,ds = \frac{1}{2\pi}\sum\int_{\bigcirc}\frac{dV}{dn}\,ds = \sum{}' m_i = M,$$

where Σ extends over all the masses and Σ' over all the masses within the circuit (none being on the circuit), gives the total mass M within the circuit.

127. Some critical comments. In the discussion of line integrals and in the future discussion of double integrals it is necessary to speak frequently of curves. For the usual problem the intuitive conception of a curve suffices. A curve as ordinarily conceived is continuous, has a continuously turning tangent line except perhaps at a finite number of angular points, and is cut by a line parallel to any given direction in only a finite number of points, except as a portion of the curve may coincide with such a line. The ideas of length and area are also applicable. For those, however, who are interested in more than the intuitive presentation of the idea of a curve and some of the matters therewith connected, the following sections are offered.

If $\phi(t)$ and $\psi(t)$ are two single valued real functions of the real variable t defined for all values in the interval $t_0 \leqq t \leqq t_1$, the pair of equations

$$x = \phi(t), \qquad y = \psi(t), \qquad t_0 \leqq t \leqq t_1, \tag{27}$$

will be said to define a *curve*. If ϕ and ψ are continuous functions of t, the curve will be called continuous. If $\phi(t_1) = \phi(t_0)$ and $\psi(t_1) = \psi(t_0)$, so that the initial and end points of the curve coincide, the curve will be called a *closed* curve provided it is continuous. If there is no other pair of values t and t' which make both $\phi(t) = \phi(t')$ and $\psi(t) = \psi(t')$, the curve will be called *simple*: in ordinary language, the curve does not cut itself. If t describes the interval from t_0 to t_1 continuously and constantly in the same sense, the point (x, y) will be said to describe the curve in a given sense; the opposite sense can be had by allowing t to describe the interval in the opposite direction.

Let the interval $t_0 \leqq t \leqq t_1$ be divided into any number n of subintervals $\Delta_1 t, \Delta_2 t, \cdots, \Delta_n t$. There will be n corresponding increments for x and y,

$$\Delta_1 x, \Delta_2 x, \cdots, \Delta_n x, \quad \text{and} \quad \Delta_1 y, \Delta_2 y, \cdots, \Delta_n y.$$

Then $\Delta_i c = \sqrt{(\Delta_i x)^2 + (\Delta_i y)^2} \leqq |\Delta_i x| + |\Delta_i y|, \quad |\Delta_i x| \leqq \Delta_i c, \quad |\Delta_i y| \leqq \Delta_i c$

are obvious inequalities. It will be necessary to consider the three sums

$$\sigma_1 = \sum_1^n |\Delta_i x|, \quad \sigma_2 = \sum_1^n |\Delta_i y|, \quad \sigma_3 = \sum_1^n \Delta_i c = \sum_1^n \sqrt{(\Delta_i x)^2 + (\Delta_i y)^2}.$$

For any division of the interval from t_0 to t_1 each of these sums has a definite positive value. When all possible modes of division are considered for any and every value of n, the sums σ_1 will form an infinite set of numbers which may be either limited or unlimited above (§ 22). In case the set is limited, the upper frontier of the set is called the *variation of x* over the curve and the curve is said to be of *limited variation in x;* in case the set is unlimited, the curve is of unlimited variation in x. Similar observations for the sums σ_2. It may be remarked that the geometric conception corresponding to the variation in x is the sum of the projections of the curve on the x-axis when the sum is evaluated arithmetically and not algebraically. Thus the variation in y for the curve $y = \sin x$ from 0 to 2π is 4. The curve $y = \sin(1/x)$ between these same limits is of unlimited variation in y. In both cases the variation in x is 2π.

If both the sums σ_1 and σ_2 have upper frontiers L_1 and L_2, the sum σ_3 will have an upper frontier $L_3 \leqq L_1 + L_2$; and conversely if σ_3 has an upper frontier, both σ_1 and σ_2 will have upper frontiers. If a new point of division is intercalated in $\Delta_i t$, the sum σ_1 cannot decrease and, moreover, it cannot increase by more than twice the oscillation of x in the interval $\Delta_i t$. For if $\Delta_{1i} x + \Delta_{2i} x = \Delta_i x$, then

$$|\Delta_{1i} x| + |\Delta_{2i} x| \geqq |\Delta_i x|, \qquad |\Delta_{1i} x| + |\Delta_{2i} x| \leqq 2(M_i - m_i).$$

Here $\Delta_{1i} t$ and $\Delta_{2i} t$ are the two intervals into which $\Delta_i t$ is divided, and $M_i - m_i$ is the oscillation in the interval $\Delta_i t$. A similar theorem is true for σ_2. It now remains to show that if the interval from t_0 to t_1 is divided sufficiently fine, the sums σ_1 and σ_2 will differ by as little as desired from their frontiers L_1 and L_2. The proof is like that of the similar problem of § 28. First, the fact that L_1 is the frontier of σ_1 shows that some method of division can be found so that $L_1 - \sigma_1 < \frac{1}{2}\epsilon$. Suppose the number of points of division is n. Let it next be assumed that $\phi(t)$ is continuous; it must then be uniformly continuous (§ 25), and hence it is possible to find a δ so small that when $\Delta_i t < \delta$ the oscillation of x is $M_i - m_i < \epsilon/4n$. Consider then any method of division for which $\Delta_i t < \delta$, and its sum σ_1'. The superposition of the former division with n points upon this gives a sum $\sigma_1'' \geqq \sigma_1'$. But $\sigma_1'' - \sigma_1' < 2n\epsilon/4n = \frac{1}{2}\epsilon$, and $\sigma_1'' \geqq \sigma_1$. Hence $L_1 - \sigma_1'' < \frac{1}{2}\epsilon$ and $L_1 - \sigma_1' < \epsilon$. A similar demonstration may be given for σ_2 and L_2.

To treat the sum σ_3 and its upper frontier L_3 note that here, too, the intercalation of an additional point of division cannot decrease σ_3 and, as

$$\sqrt{(\Delta x)^2 + (\Delta y)^2} \leqq |\Delta x| + |\Delta y|,$$

it cannot increase σ_3 by more than twice the sum of the oscillations of x and y in the interval Δt. Hence if the curve is continuous, that is, if both x and y are continuous, the division of the interval from t_0 to t_1 can be taken so fine that σ_3 shall

differ from its upper frontier L_3 by less than any assigned quantity, no matter how small. In this case $L_3 = s$ is called the *length of the curve*. It is therefore seen that *the necessary and sufficient condition that any continuous curve shall have a length is that its Cartesian coördinates x and y shall both be of limited variation.* It is clear that if the frontiers $L_1(t)$, $L_2(t)$, $L_3(t)$ from t_0 to any value of t be regarded as functions of t, they are continuous and nondecreasing functions of t, and that $L_3(t)$ is an increasing function of t; it would therefore be possible to take s in place of t as the parameter for any continuous curve having a length. Moreover if the derivatives x' and y' of x and y with respect to t exist and are continuous, the derivative s' exists, is continuous, and is given by the usual formula $s' = \sqrt{x'^2 + y'^2}$. This will be left as an exercise; so will the extension of these considerations to three dimensions or more.

In the sum $x_1 - x_0 = \Sigma\Delta_i x$ of the actual, not absolute, values of $\Delta_i x$ there may be both positive and negative terms. Let π be the sum of the positive terms and ν be the sum of the negative terms. Then

$$x_1 - x_0 = \pi - \nu, \qquad \sigma_1 = \pi + \nu, \qquad 2\pi = x_1 - x_0 + \sigma_1, \qquad 2\nu = x_0 - x_1 + \sigma_1.$$

As σ_1 has an upper frontier L_1 when x is of limited variation, and as x_0 and x_1 are constants, the sums π and ν have upper frontiers. Let these be Π and N. Considered as functions of t, neither $\Pi(t)$ nor $N(t)$ can decrease. Write $x(t) = x_0 + \Pi(t) - N(t)$. Then the function $x(t)$ of limited variation has been resolved into the difference of two functions each of limited variation and nondecreasing. As a limited nondecreasing function is integrable (Ex. 7, p. 54), this shows that *a function is integrable over any interval over which it is of limited variation.* That the difference $x = x'' - x'$ of two limited and nondecreasing functions must be a function of limited variation follows from the fact that $|\Delta x| \leqq |\Delta x''| + |\Delta x'|$. Furthermore if

$$x = x_0 + \Pi - N \quad \text{be written} \quad x = [x_0 + \Pi + |x_0| + t - t_0] - [N + |x_0| + t - t_0],$$

it is seen that *a function of limited variation can be regarded as the difference of two positive functions which are constantly increasing, and that these functions are continuous if the given function $x(t)$ is continuous.*

Let the curve C defined by the equations $x = \phi(t)$, $y = \psi(t)$, $t_0 \leqq t \leqq t_1$, be continuous. Let $P(x, y)$ be a continuous function of (x, y). Form the sum

$$\sum P(\xi_i, \eta_i)\,\Delta_i x = \sum P(\xi_i, \eta_i)\,\Delta_i x'' - \sum P(\xi_i, \eta_i)\,\Delta_i x', \tag{28}$$

where $\Delta_1 x, \Delta_2 x, \cdots$ are the increments corresponding to $\Delta_1 t, \Delta_2 t, \cdots$, where (ξ_i, η_i) is the point on the curve which corresponds to some value of t in $\Delta_i t$, where x is assumed to be of limited variation, and where x'' and x' are two continuous increasing functions whose difference is x. As x'' (or x') is a continuous and constantly increasing function of t, it is true inversely (Ex. 10, p. 45) that t is a continuous and constantly increasing function of x'' (or x'). As $P(x, y)$ is continuous in (x, y), it is continuous in t and also in x'' and x'. Now let $\Delta_i t \doteq 0$; then $\Delta_i x'' \doteq 0$ and $\Delta_i x' \doteq 0$. Also

$$\lim \sum P_i \Delta_i x'' = \int_{x''_0}^{x''_1} P\,dx'' \quad \text{and} \quad \lim \sum P_i \Delta_i x' = \int_{x'_0}^{x'_1} P\,dx'.$$

The limits exist and are integrals simply because P is continuous in x'' or in x'. Hence *the sum on the left of* (28) *has a limit and*

$$\lim \sum P\Delta_i x = \int_{C,\,x_0}^{x_1} P\,dx = \int_{x''_0}^{x''_1} P\,dx'' - \int_{x'_0}^{x'_1} P\,dx'$$

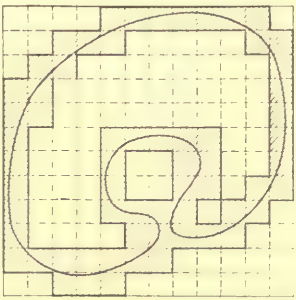
may be defined as the line integral of P along the curve C of limited variation in x. The assumption that y is of limited variation and that $Q(x, y)$ is continuous would lead to a corresponding line integral. *The assumption that both x and y are of limited variation, that is, that the curve is rectifiable, and that P and Q are continuous would lead to the existence of the line integral*

$$\int_{C\,x_0, y_0}^{x_1, y_1} P(x, y)\,dx + Q(x, y)\,dy.$$

A considerable theory of line integrals over general rectifiable curves may be constructed. The subject will not be carried further at this point.

128. The question of *the area of a curve* requires careful consideration. In the first place note that the intuitive closed plane curve which does cut itself is intuitively believed to divide the plane into two regions, one interior, one exterior to the curve; and these regions have the property that any two points of the same region may be connected by a continuous curve which does not cut the given curve, whereas any continuous curve which connects any point of one region to a point of the other must cut the given curve. The first question which arises with regard to the general closed simple curve of page 308 is: Does such a curve divide the plane into just two regions with the properties indicated, that is, is there an interior and exterior to the curve? *The answer is affirmative*, but the proof is somewhat difficult — not because the statement of the problem is involved or the proof replete with advanced mathematics, but rather because the statement is so simple and elementary that there is little to work with and the proof therefore requires the keenest and most tedious logical analysis. The theorem that a closed simple plane curve has an interior and an exterior will therefore be assumed.

As the functions $x(t)$, $y(t)$ which define the curve are continuous, they are limited, and it is possible to draw a rectangle with sides $x = a$, $x = b$, $y = c$, $y = d$ so as entirely to surround the curve. This rectangle may next be ruled with a number of lines parallel to its sides, and thus be divided into smaller rectangles. These little rectangles may be divided into three categories, those outside the curve, those inside the curve, and those upon the curve. By one upon the curve is meant one which has so much as a single point of its perimeter or interior upon the curve. Let A, A_i, A_u, A_e denote the area of the large rectangle, the sum of the areas of the small rectangles, which are interior to the curve, the sum of the areas of those upon the curve, and the sum of those exterior to it. Of course $A = A_i + A_u + A_e$. Now if all methods of ruling be considered, the quantities A_i will have an upper frontier L_i, the quantities A_e will have an upper frontier L_e, and the quantities A_u will have a lower frontier l_u. If to any method of ruling new rulings be added, the quantities A_i and A_e become A_i' and A_e' with the conditions $A_i' \geqq A_i$, $A_e' \geqq A_e$, and hence $A_u' \leqq A_u$. From this it follows that $A = L_i + l_u + L_e$. For let there be three modes of ruling which for the respective cases A_i, A_e, A_u make these three quantities differ from their frontiers L_i, L_e, l_u by less than $\frac{1}{3}\epsilon$. Then the superposition of the three systems of rulings gives rise to a ruling for which A_i', A_e', A_u' must differ from the frontier values by less than



$\frac{1}{3}\epsilon$, and hence the sum $L_i + l_u + L_e$, which is constant, differs from the constant A by less than ϵ, and must therefore be equal to it.

It is now possible to *define as the* (*qualified*) *areas of the curve*

$$L_i = \text{inner area}, \qquad l_u = \text{area on the curve}, \qquad L_i + l_u = \text{total area}.$$

In the case of curves of the sort intuitively familiar, the limit l_u is zero and $L_i = A - L_e$ becomes merely the (unqualified) area bounded by the curve. The question arises: Does the same hold for the general curve here under discussion? This time *the answer is negative;* for there are curves which, though closed and simple, are still so sinuous and meandering that a finite area l_u lies upon the curve, that is, there is a finite area so bestudded with points of the curve that no part of it is free from points of the curve. This fact again will be left as a statement without proof. Two further facts may be mentioned.

In the first place there is applicable a theorem like Theorem 21, p. 51, namely: It is possible to find a number δ so small that, when the intervals between the rulings (both sets) are less than δ, the sums A_u, A_i, A_e differ from their frontiers by less than 2ϵ. For there is, as seen above, some method of ruling such that these sums differ from their frontiers by less than ϵ. Moreover, the adding of a single new ruling cannot change the sums by more than ΔD, where Δ is the largest interval and D the largest dimension of the rectangle. Hence if the total number of intervals (both sets) for the given method is N and if δ be taken less than $\epsilon/N\Delta D$, the ruling obtained by superposing the given ruling upon a ruling where the intervals are less than δ will be such that the sums differ from the given ones by less than ϵ, and hence the ruling with intervals less than δ can only give rise to sums which differ from their frontiers by less than 2ϵ.

In the second place it should be observed that the limits L_i, l_u have been obtained by means of all possible modes of ruling where the rules were parallel to the x- and y-axes, and that there is no *a priori* assurance that these same limits would have been obtained by rulings parallel to two other lines of the plane or by covering the plane with a network of triangles or hexagons or other figures. In any thorough treatment of the subject of area such matters would have to be discussed. That the discussion is not given here is due entirely to the fact that these critical comments are given not so much with the desire to establish certain theorems as with the aim of showing the reader the sort of questions which come up for consideration in the rigorous treatment of such elementary matters as "the area of a plane curve," which he may have thought he "knew all about."

It is a common intuitive conviction that if a region like that formed by a square be divided into two regions by a continuous curve which runs across the square from one point of the boundary to another, the area of the square and the sum of the areas of the two parts into which it is divided are equal, that is, the curve (counted twice) and the two portions of the perimeter of the square form two simple closed curves, and it is expected that the sum of the areas of the curves is the area of the square. Now in case the curve is such that the frontiers l_u and l'_u formed for the two curves are not zero, it is clear that the sum $L_i + L'_i$ for the two curves will not give the area of the square but a smaller area, whereas the sum $(L_i + l_u) + (L'_i + l'_u)$ will give a greater area. Moreover in this case, it is not easy to formulate a general definition of area applicable to each of the regions and such that the sum of the areas shall be equal to the area of the combined region. But if l_u and l'_u both vanish, then the sum $L_i + L'_i$ does give the combined area.

It is therefore customary to *restrict the application of the term "area" to such simple closed curves as have $l_u = 0$, and to say that the quadrature of such curves is possible*, but that the quadrature of curves for which $l_u \neq 0$ is impossible.

It may be proved that: *If a curve is rectifiable or even if one of the functions $x(t)$ or $y(t)$ is of limited variation, the limit l_u is zero and the quadrature of the curve is possible.* For let the interval $t_0 \leqq t \leqq t_1$ be divided into intervals $\Delta_1 t, \Delta_2 t, \cdots$ in which the oscillations of x and y are $\epsilon_1, \epsilon_2, \cdots, \eta_1, \eta_2, \cdots$. Then the portion of the curve due to the interval $\Delta_i t$ may be inscribed in a rectangle $\epsilon_i \eta_i$, and that portion of the curve will lie wholly within a rectangle $2\epsilon_i \cdot 2\eta_i$ concentric with this one. In this way may be obtained a set of rectangles which entirely contain the curve. The total area of these rectangles must exceed l_u. For if all the sides of all the rectangles be produced so as to rule the plane, the rectangles which go to make up A_u for this ruling must be contained within the original rectangles, and as $A_u > l_u$, the total area of the original rectangles is greater than l_u. Next suppose $x(t)$ is of limited variation and is written as $x_0 + \Pi(t) - N(t)$, the difference of two nondecreasing functions. Then $\Sigma\epsilon_i \leqq \Pi(t_1) + N(t_1)$, that is, the sum of the oscillations of x cannot exceed the total variation of x. On the other hand as $y(t)$ is continuous, the divisions $\Delta_i t$ could have been taken so small that $\eta_i < \eta$. Hence

$$l_u < A_u \leqq \sum 2\epsilon_i \cdot 2\eta_i < 4\eta \sum \epsilon_i \leqq 4\eta [\Pi(t_1) + N(t_1)].$$

The quantity may be made as small as desired, since it is the product of a finite quantity by η. Hence $l_u = 0$ and the quadrature is possible.

It may be observed that if $x(t)$ or $y(t)$ or both are of limited variation, one or all of the three curvilinear integrals

$$-\int y dx, \qquad \int x dy, \qquad \tfrac{1}{2}\int x dy - y dx$$

may be defined, and that it should be expected that in this case the value of the integral or integrals would give the area of the curve. In fact if one desired to deal only with rectifiable curves, it would be possible to take one or all of these integrals as the *definition* of area, and thus to obviate the discussions of the present article. It seems, however, advisable at least to point out the problem of quadrature in all its generality, especially as the treatment of the problem is very similar to that usually adopted for double integrals (§ 132). From the present viewpoint, therefore, it would be a proposition for demonstration that the curvilinear integrals in the cases where they are applicable do give the value of the area as here defined, but the demonstration will not be undertaken.

EXERCISES

1. For the continuous curve (27) prove the following properties:

(α) Lines $x = a$, $x = b$ may be drawn such that the curve lies entirely between them, has at least one point on each line, and cuts every line $x = \xi$, $a < \xi < b$, in at least one point; similarly for y.

(β) From $p = x \cos\alpha + y \sin\alpha$, the normal equation of a line, prove the propositions like those of (α) for lines parallel to any direction.

(γ) If (ξ, η) is any point of the xy-plane, show that the distance of (ξ, η) from the curve has a minimum and a maximum value.

(δ) If $m(\xi, \eta)$ and $M(\xi, \eta)$ are the minimum and maximum distances of (ξ, η) from the curve, the functions $m(\xi, \eta)$ and $M(\xi, \eta)$ are continuous functions of (ξ, η). Are the coördinates $x(\xi, \eta)$, $y(\xi, \eta)$ of the points on the curve which are at minimum (or maximum) distance from (ξ, η) continuous functions of (ξ, η)?

(ϵ) If $t', t'', \cdots, t^{(k)}, \cdots$ are an infinite set of values of t in the interval $t_0 \leqq t \leqq t_1$ and if t^0 is a point of condensation of the set, then $x^0 = \phi(t^0)$, $y^0 = \psi(t^0)$ is a point of condensation of the set of points (x', y'), (x'', y''), $\cdots$, $(x^{(k)}, y^{(k)})$, $\cdots$ corresponding to the set of values $t', t'' \cdots, t^{(k)}, \cdots$.

(ζ) Conversely to (ϵ) show that if (x', y'), (x'', y''), $\cdots$, $(x^{(k)}, y^{(k)})$, $\cdots$ are an infinite set of points on the curve and have a point of condensation (x^0, y^0), then the point (x^0, y^0) is also on the curve.

(η) From (ζ) show that if a line $x = \xi$ cuts the curve in a set of points $y', y'', \cdots$, then this suite of y's contains its upper and lower frontiers and has a maximum or minimum.

2. Define and discuss rectifiable curves in space.

3. Are $y = x^2 \sin \frac{1}{x}$ and $y = \sqrt{x} \sin \frac{1}{x}$ rectifiable between $x = 0$, $x = 1$?

4. If $x(t)$ in (27) is of total variation $\mathrm{П}(t_1) + \mathrm{N}(t_1)$, show that

$$\int_{C x_0}^{x_1} P(x, y)\, dx < M[\mathrm{П}(t_1) + \mathrm{N}(t_1)],$$

where M is the maximum value of $P(x, y)$ on the curve.

5. Consider the function $\theta(\xi, \eta, t) = \tan^{-1} \dfrac{\eta - y(t)}{\xi - x(t)}$ which is the inclination of the line joining a point (ξ, η) not on the curve to a point (x, y) on the curve. With the notations of Ex. 1 (δ) show that

$$|\Delta_t \theta| = |\theta(\xi, \eta, t + \Delta t) - \theta(\xi, \eta, t)| < \frac{2M\delta}{m - 2M\delta},$$

where $\delta > |\Delta x|$ and $\delta > |\Delta y|$ may be made as small as desired by taking Δt sufficiently small and where it is assumed that $m \neq 0$.

6. From Ex. 5 infer that $\theta(\xi, \eta, t)$ is of limited variation when t describes the interval $t_0 \leqq t \leqq t_1$ defining the curve. Show that $\theta(\xi, \eta, t)$ is continuous in (ξ, η) through any region for which $m > 0$.

7. Let the parameter t vary from t_0 to t_1 and suppose the curve (27) is closed so that (x, y) returns to its initial value. Show that the initial and final values of $\theta(\xi, \eta, t)$ differ by an integral multiple of 2π. Hence infer that this difference is constant over any region for which $m > 0$. In particular show that the constant is 0 over all distant regions of the plane. It may be remarked that, by the study of this change of θ as t describes the curve, a proof may be given of the theorem that the closed continuous curve divides the plane into two regions, one interior, one exterior.

8. Extend the last theorem of § 123 to rectifiable curves.

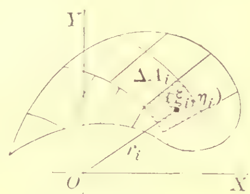
CHAPTER XII

ON MULTIPLE INTEGRALS

129. Double sums and double integrals. Suppose that a body of matter is so thin and flat that it can be considered to lie in a plane. If any small portion of the body surrounding a given point $P(x, y)$ be considered, and if its mass be denoted by Δm and its area by ΔA, the average (surface) density of the portion is the quotient $\Delta m/\Delta A$, and the actual density at the point P is defined as the limit of this quotient when $\Delta A \doteq 0$, that is,

$$D(x, y) = \lim_{\Delta A \doteq 0} \frac{\Delta m}{\Delta A}.$$

The density may vary from point to point. Now conversely suppose that the density $D(x, y)$ of the body is a known function of (x, y) and that it be required to find the total mass of the body. Let the body be considered as divided up into a large number of pieces each of which is *small in every direction*, and let ΔA_i be the area of any piece. If (ξ_i, η_i) be any point in ΔA_i, the density at that point is $D(\xi_i, \eta_i)$ and the amount of matter in the piece is approximately $D(\xi_i, \eta_i)\Delta A_i$ provided the density be regarded as continuous, that is, as not varying much over so small an area. Then the sum



$$D(\xi_1, \eta_1)\Delta A_1 + D(\xi_2, \eta_2)\Delta A_2 + \cdots + D(\xi_n, \eta_n)\Delta A_n = \sum D(\xi_i, \eta_i)\Delta A_i,$$

extended over all the pieces, is an approximation to the total mass, and may be sufficient for practical purposes if the pieces be taken tolerably small.

The process of dividing a body up into a large number of small pieces of which it is regarded as the sum is a device often resorted to: for the properties of the small pieces may be known approximately, so that the corresponding property for the whole body can be obtained approximately by summation. Thus by definition the moment of inertia of a small particle of matter relative to an axis is mr^2, where m is the mass of the particle and r its distance from the axis. If therefore the moment of inertia of a plane body with respect to an axis perpendicular

to its plane were required, the body would be divided into a large number of small portions as above. The mass of each portion would be approximately $D(\xi_i, \eta_i)\Delta A_i$ and the distance of the portion from the axis might be considered as approximately the distance r_i from the point where the axis cut the plane to the point (ξ_i, η_i) in the portion. The moment of inertia would be

$$D(\xi_1, \eta_1)r_1^2\Delta A_1 + \cdots + D(\xi_n, \eta_n)r_n^2\Delta A_n = \sum D(\xi_i, \eta_i)r_i^2\Delta A_i,$$

or nearly this, where the sum is extended over all the pieces.

These sums may be called *double* sums because they extend over two dimensions. To pass from the approximate to the actual values of the mass or moment of inertia or whatever else might be desired, the underlying idea of a division into parts and a subsequent summation is kept, but there is added to this the idea of passing to a limit. Compare §§ 16–17. Thus

$$\underset{n=\infty,\ \Delta A_i \doteq 0}{\text{limit}} \sum D(\xi_i, \eta_i)\Delta A_i \quad \text{and} \quad \underset{n=\infty,\ \Delta A_i \doteq 0}{\text{limit}} \sum D(\xi_i, \eta_i)r_i^2\Delta A_i$$

would be taken as the total mass or inertia, where the sum over n divisions is replaced by the limit of that sum as the number of divisions becomes infinite and each becomes small in every direction. The limits are indicated by a sign of integration, as

$$\lim \sum D(\xi_i, \eta_i)\Delta A_i = \int D(x, y)\,dA, \qquad \lim \sum D(\xi_i, \eta_i)r_i^2\Delta A_i = \int Dr^2 dA.$$

The use of the limit is of course dependent on the fact that the limit is actually approached, and for practical purposes it is further dependent on the invention of some way of evaluating the limit. Both these questions have been treated when the sum is a simple sum (§§ 16–17, 28–30, 35); they must now be treated for the case of a double sum like those above.

130. Consider again the problem of finding the mass and let D_i be used briefly for $D(\xi_i, \eta_i)$. Let M_i be the maximum value of the density in the piece ΔA_i and let m_i be the minimum value. Then

$$m_i\Delta A_i \leqq D_i\Delta A_i \leqq M_i\Delta A_i.$$

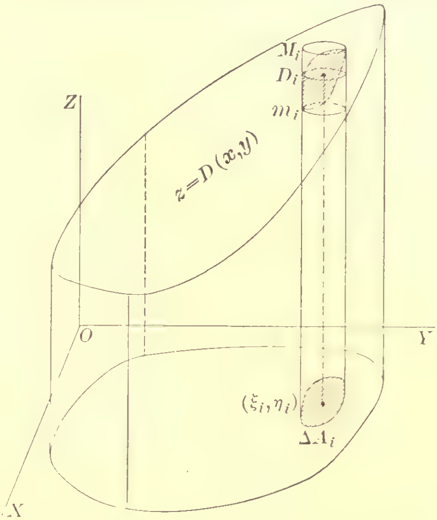
In this way any approximate expression $D_i\Delta A_i$ for the mass is shut in between two values, of which one is surely not greater than the true mass and the other surely not less. Form the sums

$$s = \sum m_i\Delta A_i \leqq \sum D_i\Delta A_i \leqq \sum M_i\Delta A_i = S$$

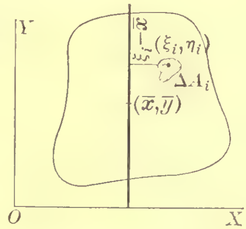
extended over all the elements ΔA_i. Now if the sums s and S approach the same limit when $\Delta A_i \doteq 0$, the sum $\Sigma D_i\Delta A_i$ which is constantly

included between s and S must also approach that limit independently of how the points (ξ_i, η_i) are chosen in the areas ΔA_i.

That s and S do approach a common limit in the usual case of a continuous function $D(x, y)$ may be shown strikingly if the surface $z = D(x, y)$ be drawn. The term $D_i \Delta A_i$ is then represented by the volume of a small cylinder upon the base ΔA_i and with an altitude equal to the height of the surface $z = D(x, y)$ above some point of ΔA_i. The sum $\Sigma D_i \Delta A_i$ of all these cylinders will be approximately the volume under the surface $z = D(x, y)$ and over the total area $A = \Sigma \Delta A_i$. The term $M_i \Delta A_i$ is represented by the volume of a small cylinder upon the base ΔA_i and circumscribed about the surface; the term $m_i \Delta A_i$, by a cylinder inscribed in the surface. When the number of elements ΔA_i is increased without limit so that each becomes indefinitely small, the three sums s, S, and $\Sigma D_i \Delta A_i$ all approach as their limit the volume under the surface and over the area A. Thus the notion of volume does for the double sum the same service as the notion of area for a simple sum.



Let the notion of the integral be applied to find *the formula for the center of gravity of a plane lamina.* Assume that the rectangular coördinates of the center of gravity are $(\bar{x}, \bar{y})$. Consider the body as divided into small areas ΔA_i. If (ξ_i, η_i) is any point in the area ΔA_i, the approximate moment of the approximate mass $D_i \Delta A_i$ in that area with respect to the line $x = \bar{x}$ is the product $(\xi_i - \bar{x}) D_i \Delta A_i$ of the mass by its distance from the line. The total exact moment would therefore be



$$\lim \sum (\xi_i - \bar{x}) D_i \Delta A_i = \int (x - \bar{x}) D(x, y) dA = 0,$$

and must vanish if the center of gravity lies on the line $x = \bar{x}$ as assumed. Then

$$\int x D(x, y) dA - \int \bar{x} D(x, y) dA = 0 \quad \text{or} \quad \int x D dA = \bar{x} \int D(x, y) dA.$$

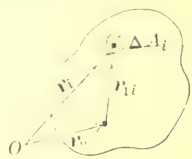
These formal operations presuppose the facts that the difference of two integrals is the integral of the difference and that the integral of a constant $\bar{x}$ times a function D

is the product of the constant by the integral of the function. It should be immediately apparent that as these rules are applicable to sums, they must be applicable to the limits of the sums. The equation may now be solved for $\bar{x}$. Then

$$\bar{x} = \frac{\int xDdA}{\int DdA} = \frac{\int xdm}{m}, \qquad \bar{y} = \frac{\int yDdA}{\int DdA} = \frac{\int ydm}{m}, \tag{1}$$

where m stands for the mass of the body and dm for DdA, just as Δm_i might replace $D_i\Delta A_i$; the result for $\bar{y}$ may be written down from symmetry.

As another example let *the kinetic energy of a lamina moving in its plane* be calculated. The use of vectors is advantageous. Let $\mathbf{r}_0$ be the vector from a fixed origin to a point which is fixed in the body, and let $\mathbf{r}_1$ be the vector from this point to any other point of the body so that



$$\mathbf{r}_i = \mathbf{r}_0 + \mathbf{r}_{1i}, \qquad \frac{d\mathbf{r}_i}{dt} = \frac{d\mathbf{r}_0}{dt} + \frac{d\mathbf{r}_{1i}}{dt} \quad \text{or} \quad \mathbf{v}_i = \mathbf{v}_0 + \mathbf{v}_{1i}.$$

The kinetic energy is $\Sigma\,\frac{1}{2}\,v_i^2\Delta m_i$ or better the integral of $\frac{1}{2}\,v^2dm$. Now

$$v_i^2 = \mathbf{v}_i\cdot\mathbf{v}_i = \mathbf{v}_0\cdot\mathbf{v}_0 + \mathbf{v}_{1i}\cdot\mathbf{v}_{1i} + 2\,\mathbf{v}_0\cdot\mathbf{v}_{1i} = v_0^2 + r_{1i}^2\omega^2 + 2\,\mathbf{v}_0\cdot\mathbf{v}_{1i}.$$

That $\mathbf{v}_{1i}\cdot\mathbf{v}_{1i} = r_{1i}^2\omega^2$, where $r_{1i} = |\mathbf{r}_{1i}|$ and ω is the angular velocity of the body about the point $\mathbf{r}_0$, follows from the fact that $\mathbf{r}_{1i}$ is a vector of constant length r_{1i} and hence $|d\mathbf{r}_{1i}| = r_{1i}d\theta$, where $d\theta$ is the angle that r_{1i} turns through, and consequently $\omega = d\theta/dt$. Next integrate over the body.

$$\begin{aligned}\int \tfrac{1}{2}\,v^2dm &= \int \tfrac{1}{2}\,v_0^2dm + \int \tfrac{1}{2}\,r_1^2\omega^2dm + \int \mathbf{v}_0\cdot\mathbf{v}_1dm \\ &= \tfrac{1}{2}\,v_0^2M + \tfrac{1}{2}\,\omega^2\int r_1^2dm + \mathbf{v}_0\cdot\int \mathbf{v}_1dm;\end{aligned} \tag{2}$$

for v_0^2 and ω^2 are constants relative to the integration over the body. Note that

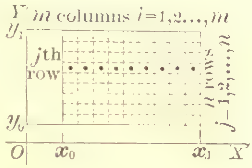
$$\mathbf{v}_0\cdot\int \mathbf{v}_1dm = 0 \quad \text{if} \quad \mathbf{v}_0 = 0 \quad \text{or if} \quad \int \mathbf{v}_1dm = \int \frac{d}{dt}\mathbf{r}_1dm = \frac{d}{dt}\int \mathbf{r}_1dm = 0.$$

But $\mathbf{v}_0 = 0$ holds only when the point $\mathbf{r}_0$ is at rest, and $\int \mathbf{r}_1dm = 0$ is the condition that $\mathbf{r}_0$ be the center of gravity. In the last case

$$T = \int \tfrac{1}{2}\,v^2dm = \tfrac{1}{2}\,v_0^2M + \tfrac{1}{2}\,\omega^2I, \qquad I = \int r_1^2dm.$$

As I is the integral which has been called the moment of inertia relative to an axis through the point $\mathbf{r}_0$ perpendicular to the plane of the body, the kinetic energy is seen to be the sum of $\frac{1}{2}\,Mv_0^2$, which would be the kinetic energy if all the mass were concentrated at the center of gravity, and of $\frac{1}{2}\,I\omega^2$, which is the kinetic energy of rotation about the center of gravity; in case $\mathbf{r}_0$ indicated a point at rest (even if only instantaneously as in § 39) the whole kinetic energy would reduce to the kinetic energy of rotation $\frac{1}{2}\,I\omega^2$. In case $\mathbf{r}_0$ indicated neither the center of gravity nor a point at rest, the third term in (2) would not vanish and the expression for the kinetic energy would be more complicated owing to the presence of this term.

131. *To evaluate the double integral in case the region is a rectangle parallel to the axes of coördinates,* let the division be made into small rectangles by drawing lines parallel to the axes. Let there be m equal divisions on one side and n on the other. There will then be mn small pieces. It will be convenient to introduce a double index and denote by ΔA_{ij} the area of the rectangle in the ith column and jth row. Let (ξ_{ij}, η_{ij}) be any point, say the middle point in the area $\Delta A_{ij} = \Delta x_i \Delta y_j$. Then the sum may be written



$$\sum_{i,j} D(\xi_{ij}, \eta_{ij})\Delta A_{ij} = D_{11}\Delta x_1\Delta y_1 + D_{21}\Delta x_2\Delta y_1 + \cdots + D_{m1}\Delta x_m\Delta y_1$$
$$+ D_{12}\Delta x_1\Delta y_2 + D_{22}\Delta x_2\Delta y_2 + \cdots + D_{m2}\Delta x_m\Delta y_2$$
$$+ \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$
$$+ D_{1n}\Delta x_1\Delta y_n + D_{2n}\Delta x_2\Delta y_n + \cdots + D_{mn}\Delta x_m\Delta y_n.$$

Now the terms in the first row are the sum of the contributions to $\Sigma_{i,j}$ of the rectangles in the first row, and so on. But

$$(D_{1j}\Delta x_1 + D_{2j}\Delta x_2 + \cdots + D_{mj}\Delta x_m)\Delta y_j = \Delta y_j \sum_i D(\xi_{ij}, \eta_{ij})\Delta x_i$$

and
$$\Delta y_j \sum_i D(\xi_{ij}, \eta_{ij})\Delta x_i = \left[\int_{x_0}^{x_1} D(x, \eta_{ij})\,dx + \zeta_j\right]\Delta y_j.$$

That is to say, by taking m sufficiently large so that the individual increments Δx_i are sufficiently small, the sum can be made to differ from the integral by as little as desired because the integral is by definition the limit of the sum. In fact

$$\zeta_j \leq \sum_i (M_{ij} - m_{ij})\Delta x_i \leq \epsilon(x_1 - x_0),$$

if ϵ be the maximum variation of $D(x, y)$ over one of the little rectangles. After thus summing up according to rows, sum up the rows. Then

$$\sum_{i,j} D_{ij}\Delta A_{ij} = \int_{x_0}^{x_1} D(x, \eta_1)\,dx\,\Delta y_1 + \int_{x_0}^{x_1} D(x, \eta_2)\,dx\,\Delta y_2$$
$$+ \cdots + \int_{x_0}^{x_1} D(x, \eta_n)\,dx\,\Delta y_n + \lambda,$$

$$\lambda = \zeta_1\Delta y_1 + \zeta_2\Delta y_2 + \cdots + \zeta_n\Delta y_n \leq \epsilon(x_1 - x_0)\sum \Delta y = \epsilon(x_1 - x_0)(y_1 - y_0).$$

If
$$\int_{x_0}^{x_1} D(x, y)\,dx = \phi(y),$$

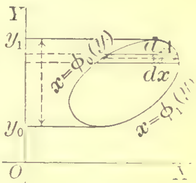
then
$$\sum_{i,j} D_{ij}\Delta A_{ij} = \phi(\eta_1)\Delta y_1 + \phi(\eta_2)\Delta y_2 + \cdots + \phi(\eta_n)\Delta y_n + \lambda$$
$$= \int_{y_0}^{y_1} \phi(y)\,dy + \kappa + \lambda, \qquad \kappa, \lambda \text{ small.}$$

Hence * $$\lim \sum_{i,j} D_{ij}\Delta A_{ij} = \int D dA = \int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx dy. \tag{3}$$

It is seen that the double integral is equal to the result obtained by first integrating with respect to x, regarding y as a parameter, and then, after substituting the limits, integrating with respect to y. If the summation had been first according to columns and second according to rows, then by symmetry

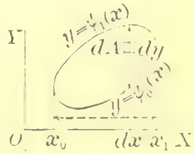
$$\int D dA = \int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx dy = \int_{x_0}^{x_1}\int_{y_0}^{y_1} D(x, y)\,dy dx. \tag{3'}$$

This is really nothing but an integration under the sign (§ 120).



If the region over which the summation is extended is not a rectangle parallel to the axes, the method could still be applied. But after summing or rather integrating according to rows, the limits would not be constants as x_0 and x_1, but would be those functions $x = \phi_0(y)$ and $x = \phi_1(y)$ of y which represent the left-hand and right-hand curves which bound the region. Thus

$$\int D dA = \int_{y_0}^{y_1}\int_{\phi_0(y)}^{\phi_1(y)} D(x, y)\,dx dy. \tag{3''}$$



And if the summation or integration had been first with respect to columns, the limits would not have been the constants y_0 and y_1, but the functions $y = \psi_0(x)$ and $y = \psi_1(x)$ which represent the lower and upper bounding curves of the region. Thus

$$\int D dA = \int_{x_0}^{x_1}\int_{\psi_0(x)}^{\psi_1(x)} D(x, y)\,dy dx. \tag{3'''}$$

The order of the integrations cannot be inverted without making the corresponding changes in the limits, the first set of limits being such functions (of the variable with regard to which the second integration is to be performed) as to sum up according to strips reaching from one side of the region to the other, and the second set of limits being constants which determine the extreme limits of the second variable so as to sum up all the strips. Although the results (3'') and (3''') are equal, it frequently happens that one of them is decidedly easier to evaluate than the other. Moreover, it has clearly been assumed that a line parallel to the

* The result may also be obtained as in Ex. 8 below.

axis of the first integration cuts the bounding curve in only two points; if this condition is not fulfilled, the area must be divided into subareas for which it is fulfilled, and the results of integrating over these smaller areas must be added algebraically to find the complete value.

To apply these rules for evaluating a double integral, consider the problem of finding the moment of inertia of a rectangle of constant density with respect to one vertex. Here

$$I = \int Dr^2 dA = D\int (x^2 + y^2)\,dA = D\int_0^b \int_0^a (x^2+y^2)\,dx\,dy$$
$$= D\int_0^b \Big[\tfrac{1}{3}x^3 + xy^2\Big]_0^a dy = D\int_0^b (\tfrac{1}{3}a^3 + ay^2)\,dy = \tfrac{1}{3}Dab(a^2+b^2).$$

If the problem had been to find the moment of inertia of an ellipse of uniform density with respect to the center, then

$$I = D\int (x^2+y^2)\,dA = D\int_{-b}^{b}\int_{-\frac{a}{b}\sqrt{b^2-y^2}}^{+\frac{a}{b}\sqrt{b^2-y^2}} (x^2+y^2)\,dx\,dy$$
$$= D\int_{-a}^{+a}\int_{-\frac{b}{a}\sqrt{a^2-x^2}}^{+\frac{b}{a}\sqrt{a^2-x^2}} (x^2+y^2)\,dx\,dy.$$

Either of these forms might be evaluated, but the moment of inertia of the whole ellipse is clearly four times that of a quadrant, and hence the simpler results

$$I = 4D\int_0^b \int_0^{\frac{a}{b}\sqrt{b^2-y^2}} (x^2+y^2)\,dx\,dy$$
$$= 4D\int_0^a \int_0^{\frac{b}{a}\sqrt{a^2-x^2}} (x^2+y^2)\,dy\,dx = \frac{\pi}{4}Dab(a^2+b^2).$$

It is highly advisable to make use of symmetry, wherever possible, to reduce the region over which the integration is extended.

132. With regard to *the more careful consideration of the limits involved in the definition of a double integral* a few observations will be sufficient. Consider the sums S and s and let $M_i\Delta A_i$ be any term of the first and $m_i\Delta A_i$ the corresponding term of the second. Suppose the area ΔA_i divided into two parts ΔA_{1i} and ΔA_{2i}, and let M_{1i}, M_{2i} be the maxima in the parts and m_{1i}, m_{2i} the minima. Then since the maximum in the whole area ΔA_i cannot be less than that in either part, and the minimum in the whole cannot be greater than that in either part, it follows that $m_{1i} \geqq m_i$, $m_{2i} \geqq m_i$, $M_{1i} \leqq M_i$, $M_{2i} \leqq M_i$, and

$$m_i\Delta A_i \leqq m_{1i}\Delta A_{1i} + m_{2i}\Delta A_{2i}, \qquad M_{1i}\Delta A_{1i} + M_{2i}\Delta A_{2i} \leqq M_i\Delta A_i.$$

Hence when one of the pieces ΔA_i is subdivided the sum S cannot increase nor the sum s decrease. Then continued inequalities may be written as

$$mA \leqq \sum m_i\Delta A_i \leqq \sum D(\xi_i, \eta_i)\Delta A_i \leqq \sum M_i\Delta A_i \leqq MA.$$

If then the original divisions ΔA_i be subdivided indefinitely, both S and s will approach limits (§§ 21–22); and if those limits are the same, the sum $\Sigma D_i\Delta A_i$ will approach that common limit as its limit independently of how the points (ξ_i, η_i) are chosen in the areas ΔA_i.

It has not been shown, however, that the limits of S and s are independent of the method of division and subdivision of the whole area. Consider therefore not only the sums S and s due to some particular mode of subdivision, but consider all such sums due to all possible modes of subdivision. As the sums S are limited below by mA they must have a lower frontier L, and as the sums s are limited above by MA they must have an upper frontier l. It must be shown that $l \leqq L$. To see this consider any pair of sums S and s corresponding to one division and any other pair of sums S' and s' corresponding to another method of division; also the sums S'' and s'' corresponding to the division obtained by combining, that is, by superposing the two methods. Now

$$S' \geqq S'' \geqq s'' \geqq s, \quad S \geqq S'' \geqq s'' \geqq s', \quad S \geqq L, \quad S' \geqq L, \quad s \leqq l, \quad s' \leqq l.$$

It therefore is seen that any S is greater than any s, whether these sums correspond to the same or to different methods of subdivision. Now if $L < l$, some S would have to be less than some s; for as L is the frontier for the sums S, there must be some such sums which differ by as little as desired from L; and in like manner there must be some sums s which differ by as little as desired from l. Hence as no S can be less than any s, the supposition $L < l$ is untrue and $L \geqq l$.

Now if for any method of division the limit of the difference

$$\lim (S - s) = \lim \sum (M_i - m_i)\,\Delta A_i = \lim \sum O_i \Delta A_i = 0$$

of the two sums corresponding to that method is zero, the frontiers L and l must be the same and both S and s approach that common value as their limit; and if the difference $S - s$ approaches zero for every method of division, the sums S and s will approach the same limit $L = l$ for all methods of division, and the sum $\Sigma D_i \Delta A_i$ will approach that limit independently of the method of division as well as independently of the selection of (ξ_i, η_i). This result follows from the fact that $L - l \leqq S - s$, $S - L \leqq S - s$, $l - s \leqq S - s$, and hence if the limit of $S - s$ is zero, then $L = l$ and S and s must approach the limit $L = l$. One case, which covers those arising in practice, in which these results are true is that in which $D(x, y)$ is continuous over the area A except perhaps upon a finite number of curves, each of which may be inclosed in a strip of area as small as desired and upon which $D(x, y)$ remains finite though it be discontinuous. For let the curves over which $D(x, y)$ is discontinuous be inclosed in strips of total area a. The contribution of these areas to the difference $S - s$ cannot exceed $(M - m)a$. Apart from these areas, the function $D(x, y)$ is continuous, and it is possible to take the divisions ΔA_i so small that the oscillation of the function over any one of them is less than an assigned number ϵ. Hence the contribution to $S - s$ is less than $\epsilon(A - a)$ for the remaining undeleted regions. The total value of $S - s$ is therefore less than $(M - m)a + \epsilon(A - a)$ and can certainly be made as small as desired.

The proof of the existence and uniqueness of the limit of $\Sigma D_i \Delta A_i$ is therefore obtained in case D is continuous over the region A except for points along a finite number of curves where it may be discontinuous provided it remains finite. Throughout the discussion the term "area" has been applied; this is justified by the previous work (§ 128). Instead of dividing the area A into elements ΔA, one may rule the area with lines parallel to the axes, as done in § 128, and consider the sums $\Sigma M \Delta x \Delta y$, $\Sigma m \Delta x \Delta y$, $\Sigma D \Delta x \Delta y$, where the first sum is extended over all the rectangles which lie within or upon the curve, where the second sum is extended over all the rectangles within the curve, and where the last extends over all rectangles

within the curve and over an arbitrary number of those upon it. In a certain sense this method is simpler, in that the area then falls out as the integral of the special function which reduces to 1 within the curve and to 0 outside the curve, and to either upon the curve. The reader who desires to follow this method through may do so for himself. It is not within the range of this book to do more in the way of rigorous analysis than to treat the simpler questions and to indicate the need of corresponding treatment for other questions.

The justification for the method of evaluating a definite double integral as given above offers some difficulties in case the function $D(x, y)$ is discontinuous. The proof of the rule may be obtained by a careful consideration of the integration of a function defined by an integral containing a parameter. Consider

$$\phi(y) = \int_{x_0}^{x_1} D(x, y)\,dx, \qquad \int_{y_0}^{y_1} \phi(y)\,dy = \int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx\,dy. \tag{4}$$

It was seen (§ 118) that $\phi(y)$ is a continuous function of y if $D(x, y)$ is a continuous function of (x, y). Suppose that $D(x, y)$ were discontinuous, but remained finite, on a finite number of curves each of which is cut by a line parallel to the x-axis in only a finite number of points. Form $\Delta\phi$ as before. Cut out the short intervals in which discontinuities may occur. As the number of such intervals is finite and as each can be taken as short as desired, their total contribution to $\phi(y)$ or $\phi(y + \Delta y)$ can be made as small as desired. For the remaining portions of the interval $x_0 \leqq x \leqq x_1$ the previous reasoning applies. Hence the difference $\Delta\phi$ can still be made as small as desired and $\phi(y)$ is continuous. If $D(x, y)$ be discontinuous along a line $y = \beta$ parallel to the x-axis, then $\phi(y)$ might not be defined and might have a discontinuity for the value $y = \beta$. But there can be only a finite number of such values if $D(x, y)$ satisfies the conditions imposed upon it in considering the double integral above. Hence $\phi(y)$ would still be integrable from y_0 to y_1. Hence

$$\int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx\,dy \qquad \text{exists}$$

and $$m(x_1 - x_0)(y_1 - y_0) \leqq \int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx\,dy \leqq M(x_1 - x_0)(y_1 - y_0)$$

under the conditions imposed for the double integral.

Now let the rectangle $x_0 \leqq x \leqq x_1$, $y_0 \leqq y \leqq y_1$ be divided up as before. Then

$$m_{ij}\Delta x_i \Delta y_j \leqq \int_y^{y+\Delta_j y}\int_x^{x+\Delta_i x} D(x, y)\,dx\,dy \leqq M_{ij}\Delta_i x \Delta_j y.$$

Add: $$\sum m_{ij}\Delta x_i \Delta y_j \leqq \sum \int_y^{y+\Delta_j y}\int_x^{x+\Delta_i x} D(x, y)\,dx\,dy \leqq \sum M_{ij}\Delta_i x \Delta_j y$$

and $$\sum \int_y^{y+\Delta_j y}\int_x^{x+\Delta_i x} D(x, y)\,dx\,dy = \int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx\,dy.$$

Now if the number of divisions is multiplied indefinitely, the limit is

$$\int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx\,dy = \lim \sum m_{ij}\Delta A_{ij} = \lim \sum M_{ij}\Delta A_{ij} = \int D(x, y)\,dA.$$

Thus the previous rule for the rectangle is proved with proper allowance for possible discontinuities. In case the area A did not form a rectangle, a rectangle could be described about it and the function $D(x, y)$ could be defined for the whole rectangle as follows: For points within A the value of $D(x, y)$ is already

defined, for points of the rectangle outside of A take $D(x, y) = 0$. The discontinuities across the boundary of A which are thus introduced are of the sort allowable for either integral in (4), and the integration when applied to the rectangle would then clearly give merely the integral over A. The limits could then be adjusted so that

$$\int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx dy = \int_{y_0}^{y_1}\int_{x=\phi_0(y)}^{x=\phi_1(y)} D(x, y)\,dx dy = \int D(x, y)\,dA.$$

The rule for evaluating the double integral by repeated integration is therefore proved.

EXERCISES

1. The sum of the moments of inertia of a plane lamina about two perpendicular lines in its plane is equal to the moment of inertia about an axis perpendicular to the plane and passing through their point of intersection.

2. The moment of inertia of a plane lamina about any point is equal to the sum of the moment of inertia about the center of gravity and the product of the total mass by the square of the distance of the point from the center of gravity.

3. If upon every line issuing from a point O of a lamina there is laid off a distance OP such that OP is inversely proportional to the square root of the moment of inertia of the lamina about the line OP, the locus of P is an ellipse with center at O.

4. Find the moments of inertia of these uniform laminas:

(α) segment of a circle about the center of the circle,
(β) rectangle about the center and about either side,
(γ) parabolic segment bounded by the latus rectum about the vertex or diameter,
(δ) right triangle about the right-angled vertex and about the hypotenuse.

5. Find by double integration the following areas:

(α) quadrantal segment of the ellipse, (β) between $y^2 = x^3$ and $y = x$,
(γ) between $3y^2 = 25x$ and $5x^2 = 9y$,
(δ) between $x^2 + y^2 - 2x = 0$, $x^2 + y^2 - 2y = 0$,
(ϵ) between $y^2 = 4ax + 4a^2$, $y^2 = -4bx + 4b^2$,
(ζ) within $(y - x - 2)^2 = 4 - x^2$,
(η) between $x^2 = 4ay$, $y(x^2 + 4a^2) = 8a^3$,
(θ) $y^2 = ax$, $x^2 + y^2 - 2ax = 0$.

6. Find the center of gravity of the areas in Ex. 5 (α), (β), (γ), (δ), and

(α) quadrant of $a^4y^2 = a^2x^4 - x^6$, (β) quadrant of $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$,
(γ) between $x^{\frac{1}{2}} = y^{\frac{1}{2}} + a^{\frac{1}{2}}$, $x + y = a$, (δ) segment of a circle.

7. Find the volumes under the surfaces and over the areas given:

(α) sphere $z = \sqrt{a^2 - x^2 - y^2}$ and square inscribed in $x^2 + y^2 = a^2$,
(β) sphere $z = \sqrt{a^2 - x^2 - y^2}$ and circle $x^2 + y^2 - ax = 0$,
(γ) cylinder $z = \sqrt{4a^2 - y^2}$ and circle $x^2 + y^2 - 2ax = 0$,
(δ) paraboloid $z = kxy$ and rectangle $0 \leqq x \leqq a$, $0 \leqq y \leqq b$,
(ϵ) paraboloid $z = kxy$ and circle $x^2 + y^2 - 2ax - 2ay = 0$,
(ζ) plane $x/a + y/b + z/c = 1$ and triangle $xy(x/a + y/b - 1) = 0$,
(η) paraboloid $z = 1 - x^2/4 - y^2/9$ above the plane $z = 0$,
(θ) paraboloid $z = (x + y)^2$ and circle $x^2 + y^2 = a^2$

8. Instead of choosing (ξ_i, η_j) as particular points, namely the middle points, of the rectangles and evaluating $\Sigma D(\xi_i, \eta_j)\Delta x_i \Delta y_j$ subject to errors λ, κ which vanish in the limit, assume the function $D(x, y)$ continuous and resolve the double integral into a double sum by repeated use of the Theorem of the Mean, as

$$\phi(y) = \int_{x_0}^{x_1} D(x, y)\,dx = \sum_i D(\xi_i, y)\,\Delta x_i, \qquad \xi\text{'s properly chosen,}$$

$$\int_{y_0}^{y_1} \phi(y)\,dy = \sum_j \phi(\eta_j)\,\Delta y_j = \sum_j \left[\sum_i D(\xi_i, \eta_j)\,\Delta x_i\right]\Delta y_j = \sum_{i,j} D(\xi_i, \eta_j)\,\Delta A_{ij}.$$

9. Consider the generalization of Osgood's Theorem (§ 35) to apply to double integrals and sums, namely: If α_{ij} are infinitesimals such that

$$\alpha_{ij} = D(\xi_i, \eta_j)\,\Delta A_{ij} + \zeta_{ij}\Delta A_{ij},$$

where ζ_{ij} is uniformly an infinitesimal, then

$$\lim \sum_{i,j} \alpha_{ij} = \int D(x, y)\,dA = \int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y)\,dx dy.$$

Discuss the statement and the result in detail in view of § 34.

10. Mark the region of the xy-plane over which the integration extends:*

$(\alpha)\ \int_0^a \int_0^x D dy dx,$ $\quad (\beta)\ \int_1^2 \int_x^{x^2} D dy dx,$ $\quad (\gamma)\ \int_0^1 \int_{y^2}^{y} D dx dy,$

$(\delta)\ \int_1^{\sqrt{2}} \int_{\frac{\sqrt{2}}{x}}^{\sqrt{3-x^2}} D dy dx,$ $\quad (\epsilon)\ \int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \int_a^{a\sqrt{2\cos 2\phi}} D dr d\phi,$ $\quad (\zeta)\ \int_a^{2a} \int_{-\frac{\pi}{6}}^{\frac{1}{2}\cos^{-1}\frac{r}{2a}} D d\phi dr.$

11. The density of a rectangle varies as the square of the distance from one vertex. Find the moment of inertia about that vertex, and about a side through the vertex.

12. Find the mass and center of gravity in Ex. 11.

13. Show that the moments of momentum (§ 80) of a lamina about the origin and about the point at the extremity of the vector $\mathbf{r}_0$ satisfy

$$\int \mathbf{r}\times\mathbf{v}dm = \mathbf{r}_0\times\int \mathbf{v}dm + \int \mathbf{r}'\times\mathbf{v}dm,$$

or the difference between the moments of momentum about P and Q is the moment about P of the total momentum considered as applied at Q.

14. Show that the formulas (1) for the center of gravity reduce to

$$\bar{x} = \frac{\int_0^a xyDdx}{\int_0^a yDdx}, \qquad \bar{y} = \frac{\int_0^a \frac{1}{2}yyDdx}{\int yDdx} \qquad \text{or} \qquad \bar{x} = \frac{\int_{x_0}^{x_1} x(y_1 - y_0)Ddx}{\int_{x_0}^{x_1}(y_1 - y_0)Ddx},$$

$$\bar{y} = \frac{\int_{x_0}^{x_1} \frac{1}{2}(y_1 + y_0)(y_1 - y_0)Ddx}{\int_{x_0}^{x_1}(y_1 - y_0)Ddx}.$$

* Exercises involving polar coördinates may be postponed until § 134 is reached, unless the student is already somewhat familiar with the subject.

when $D(x, y)$ reduces to a function $D(x)$, it being understood that for the first two the area is bounded by $x = 0$, $x = a$, $y = f(x)$, $y = 0$, and for the second two by $x = x_0$, $x = x_1$, $y_1 = f_1(x)$, $y_0 = f_0(x)$.

15. A rectangular hole is cut through a sphere, the axis of the hole being a diameter of the sphere. Find the volume cut out. Discuss the problem by double integration and also as a solid with parallel bases.

16. Show that the moment of momentum of a plane lamina about a fixed point or about the instantaneous center is $I\omega$, where ω is the angular velocity and I the moment of inertia. Is this true for the center of gravity (not necessarily fixed)? Is it true for other points of the lamina?

17. Invert the order of integration in Ex. 10 and in $\int_{-1}^{1}\int_{\sqrt{4-y^2}}^{\sqrt{3}y+2\sqrt{3}} D\,dy\,dx$.

18. In these integrals cut down the region over which the integral must be extended to the smallest possible by using symmetry, and evaluate if possible:

(α) the integral of Ex. 17 with $D = y^3 - 2x^2y$.

(β) the integral of Ex. 17 with $D = (x - 2\sqrt{3})^2 y$ or $D = (x - 2\sqrt{3})\,y^2$,

(γ) the integral of Ex. 10(ϵ) with $D = r(1 + \cos\phi)$ or $D = \sin\phi\cos\phi$.

19. The curve $y = f(x)$ between $x = a$ and $x = b$ is constantly increasing. Express the volume obtained by revolving the curve about the x-axis as $\pi[f(a)]^2(b - a)$ plus a double integral, in rectangular and in polar coördinates.

20. Express the area of the cardioid $r = a(1 - \cos\phi)$ by means of double integration in rectangular coördinates with the limits for both orders of integration.

133. Triple integrals and change of variable. In the extension from double to triple and higher integrals there is little to cause difficulty. For the discussion of the triple integral the same foundation of mass and density may be made fundamental. If $D(x, y, z)$ is the density of a body at any point, the mass of a small volume of the body surrounding the point (ξ_i, η_i, ζ_i) will be approximately $D(\xi_i, \eta_i, \zeta_i)\Delta V_i$, and will surely lie between the limits $M_i\Delta V_i$ and $m_i\Delta V_i$, where M_i and m_i are the maximum and minimum values of the density in the element of volume ΔV_i. The total mass of the body would be taken as

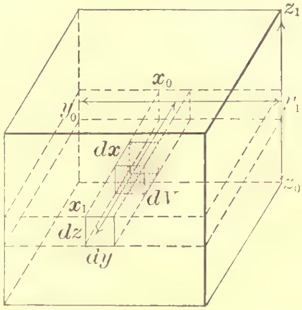
$$\lim_{\Delta V_i \doteq 0} \sum D(\xi_i, \eta_i, \zeta_i)\Delta V_i = \int D(x, y, z)\,dV, \tag{5}$$

where the sum is extended over the whole body. That the limit of the sum exists and is independent of the method of choice of the points (ξ_i, η_i, ζ_i) and of the method of division of the total volume into elements ΔV_i, provided $D(x, y, z)$ is continuous and the elements ΔV_i approach zero in such a manner that they become small in every direction, is tolerably apparent.

The evaluation of the triple integral by repeated or iterated integration is the immediate generalization of the method used for the double integral. If the region over which the integration takes place is a rectangular parallelepiped with its edges parallel to the axes, the integral is

$$\int D(x, y, z)\,dV = \int_{z_0}^{z_1}\int_{y_0}^{y_1}\int_{x_0}^{x_1} D(x, y, z)\,dx\,dy\,dz. \qquad (5')$$

The integration with respect to x adds up the mass of the elements in the column upon the base $dydz$, the integration with respect to y then adds these columns together into a lamina of thickness dz, and the integration with respect to z finally adds together the laminas and obtains the mass in the entire parallelepiped. This could be done in other orders; in fact the integration might be performed first with regard to any of the three variables, second with either of the others, and finally with the last. There are, therefore, six equivalent methods of integration.



If the region over which the integration is desired is not a rectangular parallelepiped, the only modification which must be introduced is to adjust the limits in the successive integrations so as to cover the entire region. Thus if the first integration is with respect to x and the region is bounded by a surface $x = \psi_0(y, z)$ on the side nearer the yz-plane and by a surface $x = \psi_1(y, z)$ on the remoter side, the integration

$$\int_{x=\psi_0(y,z)}^{x=\psi_1(y,z)} D(x, y, z)\,dx\,dy\,dz = \Omega(y, z)\,dy\,dz$$

will add up the mass in elements of the column which has the cross section $dydz$ and is intercepted between the two surfaces. The problem of adding up the columns is merely one in double integration over the region of the yz-plane upon which they stand; this region is the projection of the given volume upon the yz-plane. The value of the integral is then

$$\int D\,dV = \int_{z_0}^{z_1}\int_{y=\phi_0(z)}^{y=\phi_1(z)} \Omega\,dy\,dz = \int_{z_0}^{z_1}\int_{\phi_0(z)}^{\phi_1(z)}\int_{\psi_0(x,y)}^{\psi_1(x,y)} D\,dx\,dy\,dz. \quad (5'')$$

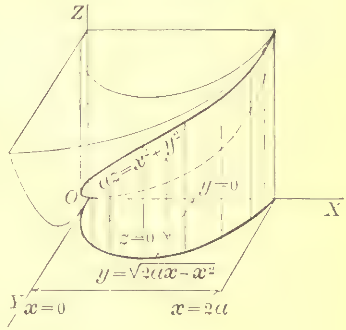
Here again the integrations may be performed in any order, provided the limits of the integrals are carefully adjusted to correspond to that order. The method may best be learned by example.

Find the mass, center of gravity, and moment of inertia about the axes of the volume of the cylinder $x^2 + y^2 - 2ax = 0$ which lies in the first octant and under paraboloid $x^2 + y^2 = az$, if the density be assumed constant. The integrals to evaluate are:

$$m = \int D dV, \quad \bar{x} = \frac{\int x dm}{m}, \quad \bar{y} = \frac{\int y dm}{m}, \quad \bar{z} = \frac{\int z dm}{m}, \tag{6}$$

$$I_x = \int D(y^2 + z^2)dV, \quad I_y = D\int (x^2 + z^2)dV, \quad I_z = D\int (x^2 + y^2)dV.$$

The consideration of how the figure looks shows that the limits for z are $z = 0$ and $z = (x^2 + y^2)/a$ if the first integration be with respect to z; then the double integral in x and y has to be evaluated over a semicircle, and the first integration is more simple if made with respect to y with limits $y = 0$ and $y = \sqrt{2ax - x^2}$, and final limits $x = 0$ and $x = 2a$ for x. If the attempt were made to integrate first with respect to y, there would be difficulty because a line parallel to the y-axis will give different limits according as it cuts both the paraboloid and cylinder or the xz-plane and cylinder; the total integral would be the sum of two integrals. There would be a similar difficulty with respect to an initial integration by x. The order of integration should therefore be z, y, x.



$$m = D\int_0^{2a}\int_0^{\sqrt{2ax-x^2}}\int_0^{(x^2+y^2)/a} dz dy dx = D\int_0^{2a}\int_0^{\sqrt{2ax-x^2}} \frac{x^2+y^2}{a} dy dx$$

$$= \frac{D}{a}\int_0^{2a}\left[x^2\sqrt{2ax - x^2} + \frac{1}{3}(2ax - x^2)^{\frac{3}{2}}\right]dx \qquad \begin{aligned} x &= a(1-\cos\theta) \\ \sqrt{2ax - x^2} &= a\sin\theta \\ dx &= a\sin\theta d\theta \end{aligned}$$

$$= Da^3\int_0^{\pi}\left[(1-\cos\theta)^2\sin^2\theta + \frac{1}{3}\sin^4\theta\right]d\theta = \frac{3}{4}\pi a^3 D$$

$$m\bar{x} = \int_0^{2a}\int_0^{\sqrt{2ax-x^2}}\int_0^{(x^2+y^2)/a} x dz dy dx = D\int_0^{2a}\int_0^{\sqrt{2ax-x^2}} \frac{x^3+xy^2}{a} dy dx$$

$$= \frac{D}{a}\int_0^{2a}\left[x^3\sqrt{2ax - x^2} + \frac{1}{3}x(2ax - x^2)^{\frac{3}{2}}\right]dx = \pi a^4 D.$$

Hence $\bar{x} = 4a/3$. The computation of the other integrals may be left as an exercise.

134. Sometimes the region over which a multiple integral is to be evaluated is such that the evaluation is relatively simple in one kind of coördinates but entirely impracticable in another kind. In addition to the rectangular coördinates the most useful systems are polar coördinates in the plane (for double integrals) and polar and cylindrical coördinates in space (for triple integrals). It has been seen (§ 40) that the element of area or of volume in these cases is

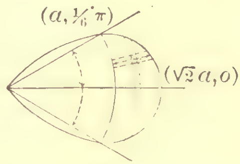
$$dA = rdrd\phi, \qquad dV = r^2\sin\theta drd\theta d\phi, \qquad dV = rdrd\phi dz. \tag{7}$$

except for infinitesimals of higher order. These quantities may be substituted in the double or triple integral and the evaluation may be made by successive integration. The proof that the substitution can be made is entirely similar to that given in §§ 34–35. The proof that the integral may still be evaluated by successive integration, with a proper choice of the limits so as to cover the region, is contained in the statement that the formal work of evaluating a multiple integral by repeated integration is independent of what the coördinates actually represent, for the reason that they could be interpreted if desired as representing rectangular coördinates.

Find the area of the part of one loop of the lemniscate $r^2 = 2a^2\cos 2\phi$ which is exterior to the circle $r = a$; also the center of gravity and the moment of inertia relative to the origin under the assumption of constant density. Here the integrals are

$$A = \int dA, \qquad A\bar{x} = \int x dA, \qquad A\bar{y} = \int y dA, \qquad I = D\int r^2 dA, \qquad m = DA.$$

The integrations may be performed first with respect to r so as to add up the elements in the little radial sectors, and then with regard to ϕ so as to add the sectors; or first with regard to ϕ so as to combine the elements of the little circular strips, and then with regard to r so as to add up the strips. Thus

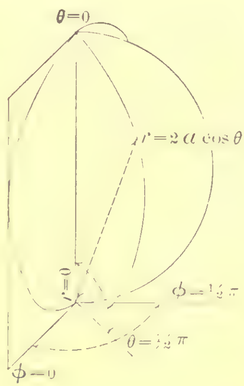


$$A = 2\int_{\phi=0}^{\frac{\pi}{6}}\int_{r=a}^{a\sqrt{2\cos 2\phi}} r dr d\phi = \int_0^{\frac{\pi}{6}}(2a^2\cos 2\phi - a^2)d\phi = \left(\frac{1}{2}\sqrt{3} - \frac{\pi}{6}\right)a^2 = .343\,a^2,$$

$$A\bar{x} = 2\int_{\phi=0}^{\frac{\pi}{6}}\int_{r=a}^{a\sqrt{2\cos 2\phi}} r\cos\phi \cdot r dr d\phi = \frac{2}{3}\int_0^{\frac{\pi}{6}}(2\sqrt{2}\,a^3\cos^{\frac{3}{2}} 2\phi - a^3)\cos\phi d\phi$$

$$= \frac{2}{3}a^3\int_0^{\frac{\pi}{6}}\left[2\sqrt{2}(1 - 2\sin^2\phi)^{\frac{3}{2}} d\sin\phi - \cos\phi d\phi\right] = \frac{\pi}{8}a^3 = .393\,a^3.$$

Hence $\bar{x} = 3\pi a/(12\sqrt{3} - 4\pi) = 1.15\,a$. The symmetry of the figure shows that $\bar{y} = 0$. The calculation of I may be left as an exercise.



Given a sphere of which the density varies as the distance from some point of the surface; required the mass and the center of gravity. If polar coördinates with the origin at the given point and the polar axis along the diameter through that point be assumed, the equation of the sphere reduces to $r = 2a\cos\theta$ where a is the radius. The center of gravity from reasons of symmetry will fall on the diameter. To cover the volume of the sphere r must vary from $r = 0$ at the origin to $r = 2a\cos\theta$ upon the sphere. The polar angle must range from $\theta = 0$ to $\theta = \frac{1}{2}\pi$, and the longitudinal angle from $\phi = 0$ to $\phi = 2\pi$. Then

$$m = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}}\int_{r=0}^{2a\cos\theta} kr\cdot r^2\sin\theta dr d\theta d\phi,$$

$$m\bar{z} = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}}\int_{r=0}^{r=2a\cos\theta} kr\cdot r\cos\theta\cdot r^2\sin\theta dr d\theta d\phi,$$

$$m = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}} 4ka^4\cos^4\theta\sin\theta d\theta d\phi = \int_0^{2\pi}\frac{4}{5}ka^4 d\phi = \frac{8\pi ka^4}{5},$$

$$m\bar{z} = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\frac{\pi}{2}}\frac{32ka^5}{5}\cos^6\theta\sin\theta d\theta d\phi = \int_0^{2\pi}\frac{32ka^5}{35}d\phi = \frac{64\pi ka^5}{35}.$$

The center of gravity is therefore $\bar{z} = 8a/7$.

Sometimes it is necessary to make a change of variable

$$x = \phi(u, v), \qquad y = \psi(u, v)$$

or

$$x = \phi(u, v, w), \qquad y = \psi(u, v, w), \qquad z = \omega(u, v, w) \tag{8}$$

in a double or a triple integral. The element of area or of volume has been seen to be (§ 63, and Ex. 7, p. 135)

$$dA = \left|J\left(\frac{x, y}{u, v}\right)\right| du dv \quad \text{or} \quad dV = \left|J\left(\frac{x, y, z}{u, v, w}\right)\right| du dv dw. \tag{8'}$$

Hence

$$\int D(x, y)dA = \int D(\phi, \psi)\left|J\left(\frac{x, y}{u, v}\right)\right| du dv \tag{8''}$$

and

$$\int D(x, y, z)dV = \int D(\phi, \psi, \omega)\left|J\left(\frac{x, y, z}{u, v, w}\right)\right| du dv dw.$$

It should be noted that the Jacobian may be either positive or negative but should not vanish; the difference between the case of positive and the case of negative values is of the same nature as the difference between an area or volume and the reflection of the area or volume. As the elements of area or volume are considered as positive when the increments of the variables are positive, the absolute value of the Jacobian is taken.

EXERCISES

1. Show that (6) are the formulas for the center of gravity of a solid body.

2. Show that $I_x = \int (y^2 + z^2)dm$, $I_y = \int (x^2 + z^2)dm$, $I_z = \int (x^2 + y^2)dm$ are the formulas for the moment of inertia of a solid about the axes.

3. Prove that the difference between the moments of inertia of a solid about any line and about a parallel line through the center of gravity is the product of the mass of the body by the square of the perpendicular distance between the lines.

4. Find the moment of inertia of a body about a line through the origin in the direction determined by the cosines l, m, n, and show that if a distance OP be laid off along this line inversely proportional to the square root of the moment of inertia, the locus of P is an ellipsoid with O as center.

5. Find the moments of inertia of these solids of uniform density:

(α) rectangular parallelepiped abc, about the edge a,
(β) ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$, about the z-axis,
(γ) circular cylinder, about a perpendicular bisector of its axis,
(δ) wedge cut from the cylinder $x^2 + y^2 = r^2$ by $z = \pm mx$, about its edge.

6. Find the volume of the solids of Ex. 5 (β), (δ), and of the:

(α) trirectangular tetrahedron between $xyz = 0$ and $x/a + y/b + z/c = 1$,
(β) solid bounded by the surfaces $y^2 + z^2 = 4ax$, $y^2 = ax$, $x = 3a$,
(γ) solid common to the two equal perpendicular cylinders $x^2 + y^2 = a^2$, $x^2 + z^2 = a^2$
(δ) octant of $\left(\frac{x}{a}\right)^{\frac{1}{2}} + \left(\frac{y}{b}\right)^{\frac{1}{2}} + \left(\frac{z}{c}\right)^{\frac{1}{2}} = 1$, ($\epsilon$) octant of $\left(\frac{x}{a}\right)^{\frac{2}{3}} + \left(\frac{y}{b}\right)^{\frac{2}{3}} + \left(\frac{z}{c}\right)^{\frac{2}{3}} = 1$.

7. Find the center of gravity in Ex. 5 (δ), Ex. 6 (α), (β), (δ), (ϵ), density uniform.

8. Find the area in these cases: (α) between $r = a \sin 2\phi$ and $r = \frac{1}{2} a$.
(β) between $r^2 = 2a^2 \cos 2\phi$ and $r = 3^{\frac{1}{4}} a$, (γ) between $r = a \sin \phi$ and $r = b \cos \phi$.
(δ) $r^2 = 2a^2 \cos 2\phi$, $r \cos \phi = \frac{1}{2}\sqrt{3}\, a$, ($\epsilon$) $r = a(1 + \cos\phi)$, $r = a$.

9. Find the moments of inertia about the pole for the cases in Ex. 8, density uniform.

10. Assuming uniform density, find the center of gravity of the area of one loop:

(α) $r^2 = 2a^2 \cos 2\phi$, (β) $r = a(1 - \cos\phi)$, (γ) $r = a \sin 2\phi$,
(δ) $r = a \sin^3 \frac{1}{3}\phi$ (small loop), (ϵ) circular sector of angle 2α.

11. Find the moments of inertia of the areas in Ex. 10 (α), (β), (γ) about the initial line.

12. If the density of a sphere decreases uniformly from D_0 at the center to D_1 at the surface, find the mass and the moment of inertia about a diameter.

13. Find the total volume of:

(α) $(x^2 + y^2 + z^2)^2 = axyz$, ($\beta$) $(x^2 + y^2 + z^2)^3 = 27a^3xyz$.

14. A spherical sector is bounded by a cone of revolution; find the center of gravity and the moment of inertia about the axis of revolution if the density varies as the nth power of the distance from the center.

15. If a cylinder of liquid rotates about the axis, the shape of the surface is a paraboloid of revolution. Find the kinetic energy.

16. Compute $J\left(\frac{x, y}{r, \phi}\right)$, $J\left(\frac{x, y, z}{r, \phi, z}\right)$, $J\left(\frac{x, y, z}{r, \phi, \theta}\right)$ and hence verify (7).

17. Sketch the region of integration and the curves $u = \text{const.}$, $v = \text{const.}$; hence show:

(α) $\int_0^c \int_{y=0}^{c-x} f(x, y)\,dxdy = \int_0^1 \int_{u=0}^{c} f(u - uv, uv)\,ududv$, if $u = y + x$, $y = uv$,

(β) $\int_0^a \int_{y=0}^{x} f(x, y)\,dxdy$

$$= \int_0^1 \int_{v=0}^{a(1+u)} f\left(\frac{v}{1+u}, \frac{uv}{1+u}\right) \frac{v}{(1+u)^2}\,dvdu \text{ if } y = xu,\ x = \frac{v}{1+u},$$

(γ) or $= \int_0^a \int_{u=0}^1 f \frac{v}{(1+u)^2}\,dudv - \int_a^{2a} \int_{u=1}^{\frac{v}{a}-1} f \frac{v}{(1+u)^2}\,dudv.$

18. Find the volume of the cylinder $r = 2a\cos\phi$ between the cone $z = r$ and the plane $z = 0$.

19. Same as Ex. 18 for cylinder $r^2 = 2a^2\cos 2\phi$; and find the moment of inertia about $r = 0$ if the density varies as the distance from $r = 0$.

20. Assuming the law of the inverse square of the distance, show that the attraction of a homogeneous sphere at a point outside the sphere is as though all the mass were concentrated at the center.

21. Find the attraction of a right circular cone for a particle at the vertex.

22. Find the attraction of (α) a solid cylinder, (β) a cylindrical shell upon a point on its axis; assume homogeneity.

23. Find the potentials, along the axes only, in Ex. 22. The potential may be defined as $\Sigma r^{-1}dm$ or as the integral of the force.

24. Obtain the formulas for the center of gravity of a sectorial area as

$$\bar{x} = \frac{1}{A}\int_{\phi_0}^{\phi_1}\frac{1}{3}r^3\cos\phi d\phi, \qquad \bar{y} = \frac{1}{A}\int_{\phi_0}^{\phi_1}\frac{1}{3}r^3\sin\phi d\phi,$$

and explain how they could be derived from the fact that the center of gravity of a uniform triangle is at the intersection of the medians.

25. Find the total illumination upon a circle of radius a, owing to a light at a distance h above the center. The illumination varies inversely as the square of the distance and directly as the cosine of the angle between the ray and the normal to the surface.

26. Write the limits for the examples worked in §§ 133 and 134 when the integrations are performed in various other orders.

27. A theorem of Pappus. If a closed plane curve be revolved about an axis which does not cut it, the volume generated is equal to the product of the area of the curve by the distance traversed by the center of gravity of the area. Prove either analytically or by infinitesimal analysis. Apply to various figures in which two of the three quantities, volume, area, position of center of gravity, are known, to find the third. Compare Ex. 3, p. 345.

135. Average values and higher integrals. The value of some special interpretation of integrals and other mathematical entities lies in the concreteness and suggestiveness which would be lacking in a purely analytical handling of the subject. For the simple integral $\int f(x)dx$ the curve $y = f(x)$ was plotted and the integral was interpreted as an area; it would have been possible to remain in one dimension by interpreting $f(x)$ as the density of a rod and the integral as the mass. In the case of the double integral $\int f(x, y)dA$ the conception of density and mass of a lamina was made fundamental; as was pointed out, it is possible to go into three dimensions and plot the surface $z = f(x, y)$

and interpret the integral as a volume. In the treatment of the triple integral $\int f(x, y, z)\,dV$ the density and mass of a body in space were made fundamental; here it would not be possible to plot $u = f(x, y, z)$ as there are only three dimensions available for plotting.

Another important interpretation of an integral is found in the conception of *average value*. If $q_1, q_2, \cdots, q_n$ are n numbers, the average of the numbers is the quotient of their sum by n.

$$\bar{q} = \frac{q_1 + q_2 + \cdots + q_n}{n} = \frac{\Sigma q_i}{n}. \tag{9}$$

If a set of numbers is formed of w_1 numbers q_1, and w_2 numbers $q_2, \cdots$, and w_n numbers q_n, so that the total number of the numbers is $w_1 + w_2 + \cdots + w_n$, the average is

$$\bar{q} = \frac{w_1 q_1 + w_2 q_2 + \cdots + w_n q_n}{w_1 + w_2 + \cdots + w_n} = \frac{\Sigma w_i q_i}{\Sigma w_i}. \tag{9'}$$

The coefficients $w_1, w_2, \cdots, w_n$, or any set of numbers which are proportional to them, are called the *weights* of $q_1, q_2, \cdots, q_n$. These definitions of average will not apply to finding the average of an infinite number of numbers because the denominator n would not be an arithmetical number. Hence it would not be possible to apply the definition to finding the average of a function $f(x)$ in an interval $x_0 \leqq x \leqq x_1$.

A slight change in the point of view will, however, lead to a definition for *the average value of a function*. Suppose that the interval $x_0 \leqq x \leqq x_1$ is divided into a number of intervals Δx_i, and that it be imagined that the number of values of $y = f(x)$ in the interval Δx_i is proportional to the length of the interval. Then the quantities Δx_i would be taken as the weights of the values $f(\xi_i)$ and the average would be

$$\bar{y} = \frac{\Sigma \Delta x_i f(\xi_i)}{\Sigma \Delta x_i}, \quad \text{or better} \quad y = \frac{\displaystyle\int_{x_0}^{x_1} f(x)\,dx}{\displaystyle\int_{x_0}^{x_1} dx} \tag{10}$$

by passing to the limit as the Δx_i's approach zero. Then

$$\bar{y} = \frac{\displaystyle\int_{x_0}^{x_1} f(x)\,dx}{x_1 - x_0} \quad \text{or} \quad \int_{x_0}^{x_1} f(x)\,dx = (x_1 - x_0)\,\bar{y}. \tag{10'}$$

In like manner if $z = f(x, y)$ be a function of two variables or $u = f(x, y, z)$ a function of three variables, the averages over an area

or volume would be defined by the integrals

$$\bar{z} = \frac{\int f(x, y)\, dA}{\int dA = A} \quad \text{and} \quad \bar{u} = \frac{\int f(x, y, z)\, dV}{\int dV = V}. \tag{10''}$$

It should be particularly noticed that *the value of the average is defined with reference to the variables of which the function averaged is a function; a change of variable will in general bring about a change in the value of the average.* For

if $\qquad y = f(x), \qquad \overline{y(x)} = \frac{1}{x_1 - x_0}\int_{x_0}^{x_1} f(x)\, dx;$

but if $\qquad y = f(\phi(t)), \qquad \overline{y(t)} = \frac{1}{t_1 - t_0}\int_{t_0}^{t_1} f(\phi(t))\, dt;$

and there is no reason for assuming that these very different expressions have the same numerical value. Thus let

$$y = x^2, \qquad 0 \leqq x \leqq 1, \qquad x = \sin t, \qquad 0 \leqq t \leqq \tfrac{1}{2}\pi,$$

$$\overline{y(x)} = \frac{1}{1}\int_0^1 x^2 dx = \frac{1}{3}, \qquad \overline{y(t)} = \frac{1}{\frac{1}{2}\pi}\int_0^{\frac{\pi}{2}} \sin^2 t dt = \frac{1}{2}.$$

The average values of x and y over a plane area are

$$\bar{x} = \frac{1}{A}\int x dA, \qquad \bar{y} = \frac{1}{A}\int y dA,$$

when the weights are taken proportional to the elements of area; but if the area be occupied by a lamina and the weights be assigned as proportional to the elements of mass, then

$$\bar{x} = \frac{1}{m}\int x dm, \qquad \bar{y} = \frac{1}{m}\int y dm,$$

and the average values of x and y are the coördinates of the center of gravity. These two averages cannot be expected to be equal unless the density is constant. The first would be called an area-average of x and y; the second, a mass-average of x and y. The mass average of the square of the distance from a point to the different points of a lamina would be

$$\overline{r^2} = k^2 = \frac{1}{M}\int r^2 dm = I/M, \tag{11}$$

and is defined as the radius of gyration of the lamina about that point; it is the quotient of the moment of inertia by the mass.

As a problem in averages consider the determination of the average value of a proper fraction; also the average value of a proper fraction subject to the condition that it be one of two proper fractions of which the sum shall be less than or equal to 1. Let x be the proper fraction. Then in the first case

$$\bar{x} = \frac{1}{1}\int_0^1 x dx = \frac{1}{2}.$$

In the second case let y be the other fraction so that $x + y \leqq 1$. Now if (x, y) be taken as coördinates in a plane, the range is over a triangle, the number of points (x, y) in the element $dxdy$ would naturally be taken as proportional to the area of the element, and the average of x over the region would be

$$\bar{x} = \frac{\int x dA}{\int dA} = \frac{\int_0^1 \int_0^{1-y} x dx dy}{\int_0^1 \int_0^{1-y} dx dy} = \frac{\int_0^1 (1 - 2y + y^2) dy}{2\int_0^1 (1-y) dy} = \frac{1}{3}.$$

Now if x were one of four proper fractions whose sum was not greater than 1, the problem would be to average x over all sets of values (x, y, z, u) subject to the relation $x + y + z + u \leqq 1$. From the analogy with the above problems, the result would be

$$\bar{x} = \lim \frac{\Sigma x \Delta x \Delta y \Delta z \Delta u}{\Sigma \Delta x \Delta y \Delta z \Delta u} = \frac{\int_{u=0}^{1} \int_{z=0}^{1-u} \int_{y=0}^{1-u-z} \int_{x=0}^{1-u-z-y} x dx dy dz du}{\int_{u=0}^{1} \int_{z=0}^{1-u} \int_{y=0}^{1-u-z} \int_{x=0}^{1-u-z-y} dx dy dz du}.$$

The evaluation of the quadruple integral gives $\bar{x} = 1/5$.

136. The foregoing problem and other problems which may arise lead to the consideration of integrals of greater multiplicity than three. It will be sufficient to mention the case of a quadruple integral. In the first place let the four variables be

$$x_0 \leqq x \leqq x_1, \qquad y_0 \leqq y \leqq y_1, \qquad z_0 \leqq z \leqq z_1, \qquad u_0 \leqq u \leqq u_1, \quad (12)$$

included in intervals with constant limits. This is analogous to the case of a rectangle or rectangular parallelepiped for double or triple integrals. The range of values of x, y, z, u in (12) may be spoken of as a rectangular volume in four dimensions, if it be desired to use geometrical as well as analytical analogy. Then the product $\Delta x_i \Delta y_i \Delta z_i \Delta u_i$ would be an element of the region. If

$$x_i \leqq \xi_i \leqq x_i + \Delta x_i, \cdots, u_i \leqq \theta_i \leqq u_i + \Delta u_i,$$

the point $(\xi_i, \eta_i, \zeta_i, \theta_i)$ would be said to lie in the element of the region. The formation of a quadruple sum

$$\sum f(\xi_i, \eta_i, \zeta_i, \theta_i) \Delta x_i \Delta y_i \Delta z_i \Delta u_i$$

could be carried out in a manner similar to that of double and triple sums, and the sum could readily be shown to have a limit when

$\Delta x_i, \Delta y_i, \Delta z_i, \Delta u_i$ approach zero, provided f is continuous. The limit of this sum could be evaluated by iterated integration

$$\lim \sum f_i \Delta x_i \Delta y_i \Delta z_i \Delta u_i = \int_{x_0}^{x_1} \int_{y_0}^{y_1} \int_{z_0}^{z_1} \int_{u_0}^{u_1} f(x, y, z, u)\, du dz dy dx$$

where the order of the integrations is immaterial.

It is possible to define regions other than by means of inequalities such as arose above. Consider

$$F(x, y, z, u) = 0 \quad \text{and} \quad F(x, y, z, u) \leqq 0,$$

where it may be assumed that when three of the four variables are given the solution of $F = 0$ gives not more than two values for the fourth. The values of x, y, z, u which make $F < 0$ are separated from those which make $F > 0$ by the values which make $F = 0$. If the sign of F is so chosen that large values of x, y, z, u make F positive, the values which give $F > 0$ will be said to be outside the region and those which give $F < 0$ will be said to be inside the region. The value of the integral of $f(x, y, z, u)$ over the region $F \leqq 0$ could be found as

$$\int_{x_0}^{x_1} \int_{y=\phi_0(x)}^{y=\phi_1(x)} \int_{z=\psi_0(x, y)}^{z=\psi_1(x, y)} \int_{u=\omega_0(x, y, z)}^{u=\omega_1(x, y, z)} f(x, y, z, u)\, du dz dy dx,$$

where $u = \omega_1(x, y, z)$ and $u = \omega_0(x, y, z)$ are the two solutions of $F = 0$ for u in terms of x, y, z, and where the triple integral remaining after the first integration must be evaluated over the range of all possible values for (x, y, z). By first solving for one of the other variables, the integrations could be arranged in another order with properly changed limits.

If a change of variable is effected such as

$$x = \phi(x', y', z', u'), \quad y = \psi(x', y', z', u'), \quad z = \chi(x', y', z', u'), \quad u = \omega(x', y', z', u') \quad (13)$$

the integrals in the new and old variables are related by

$$\iiiint f(x, y, z, u)\, dx dy dz du = \iiiint f(\phi, \psi, \chi, \omega) \left| J\left(\frac{x, y, z, u}{x', y', z', u'}\right) \right| dx' dy' dz' du'. \quad (14)$$

The result may be accepted as a fact in view of its analogy with the results (8) for the simpler cases. A proof, however, may be given which will serve equally well as another way of establishing those results, — a way which does not depend on the somewhat loose treatment of infinitesimals and may therefore be considered as more satisfactory. In the first place note that from the relation (33) of p. 134 involving Jacobians, and from its generalization to several variables, it appears that if the change (14) is possible for each of two transformations, it is possible for the succession of the two. Now for the simple transformation

$$x = x', \qquad y = y', \qquad z = z', \qquad u = \omega(x', y', z', u') = \omega(x, y, z, u'). \quad (13')$$

which involves only one variable, $J = \partial\omega/\partial u'$, and here

$$\int f(x, y, z, u)\,du = \int f(x, y, z, u')\left|\frac{\partial u}{\partial u'}\right| du' = \int f(x', y', z', u')\,\big|J\big|\,du',$$

and each side may be integrated with respect to x, y, z. Hence (14) is true in this case. For the transformation

$$x = \phi(x', y', z', u'),\quad y = \psi(x', y', z', u'),\quad z = \chi(x', y', z', u'),\quad u = u',\qquad (13'')$$

which involves only three variables, $J\left(\frac{x, y, z, u}{x', y', z', u'}\right) = J\left(\frac{x, y, z}{x', y', z'}\right)$ and

$$\iiint f(x, y, z, u)\,dxdydz = \iiint f(\phi, \psi, \chi, u)\,|J|\,dx'dy'dz',$$

and each side may be integrated with respect to u. The rule therefore holds in this case. It remains therefore merely to show that any transformation (13) may be resolved into the succession of two such as (13′), (13″). Let

$$x_1 = x',\qquad y_1 = y',\qquad z_1 = z',\qquad u_1 = \omega(x', y', z', u') = \omega(x_1, y_1, z_1, u').$$

Solve the equation $u_1 = \omega(x_1, y_1, z_1, u')$ for $u' = \omega_1(x_1, y_1, z_1, u_1)$ and write

$$x = \phi(x_1, y_1, z_1, \omega_1),\qquad y = \psi(x_1, y_1, z_1, \omega_1),\qquad z = \chi(x_1, y_1, z_1, \omega_1),\qquad u = u_1.$$

Now by virtue of the value of ω_1, this is of the type (13″), and the substitution of x_1, y_1, z_1, u_1 in it gives the original transformation.

EXERCISES

1. Determine the average values of these functions over the intervals:

(α) x^2, $0 \leqq x \leqq 10$, (β) $\sin x$, $0 \leqq x \leqq \frac{1}{2}\pi$,
(γ) x^n, $0 \leqq x \leqq n$, (δ) $\cos^n x$, $0 \leqq x \leqq \frac{1}{2}\pi$.

2. Determine the average values as indicated:

(α) ordinate in a semicircle $x^2 + y^2 = a^2$, $y > 0$, with x as variable,
(β) ordinate in a semicircle, with the arc as variable,
(γ) ordinate in semiellipse $x = a\cos\phi$, $y = b\sin\phi$, with ϕ as variable,
(δ) focal radius of ellipse, with equiangular spacing about focus,
(ϵ) focal radius of ellipse, with equal spacing along the major axis,
(ζ) chord of a circle (with the most natural assumption).

3. Find the average height of so much of these surfaces as lies above the xy-plane:

(α) $x^2 + y^2 + z^2 = a^2$, (β) $z = a^4 - p^2x^2 - q^2y^2$, (γ) $e^z = 4 - x^2 - y^2$.

4. If a man's height is the average height of a conical tent, on how much of the floor space can he stand erect?

5. Obtain the average values of the following:

(α) distance of a point in a square from the center, (β) ditto from vertex,
(γ) distance of a point in a circle from the center, (δ) ditto for sphere,
(ϵ) distance of a point in a sphere from a fixed point on the surface.

6. From the S.W. corner of a township persons start in random directions between N. and E. to walk across the township. What is their average walk? Which has it?

7. On each of the two legs of a right triangle a point is selected and the line joining them is drawn. Show that the average of the area of the square on this line is $\frac{1}{3}$ the square on the hypotenuse of the triangle.

8. A line joins two points on opposite sides of a square of side a. What is the ratio of the average square on the line to the given square?

9. Find the average value of the sum of the squares of two proper fractions. What are the results for three and for four fractions?

10. If the sum of n proper fractions cannot exceed 1, show that the average value of any one of the fractions is $1/(n+1)$.

11. The average value of the product of k proper fractions is 2^{-k}.

12. Two points are selected at random within a circle. Find the ratio of the average area of the circle described on the line joining them as diameter to the area of the circle.

13. Show that $J = r^3 \sin^2\theta \sin\phi$ for the transformation

$$x = r\cos\theta, \quad y = r\sin\theta\cos\phi, \quad z = r\sin\theta\sin\phi\cos\psi, \quad u = r\sin\theta\sin\phi\sin\psi,$$

and prove that all values of x, y, z, u defined by $x^2 + y^2 + z^2 + u^2 \leqq a^2$ are covered by the range $0 \leqq r \leqq a$, $0 \leqq \theta \leqq \pi$, $0 \leqq \phi \leqq \pi$, $0 \leqq \psi \leqq 2\pi$. What range will cover all positive values of x, y, z, u?

14. The sum of the squares of two proper fractions cannot exceed 1. Find the average value of one of the fractions.

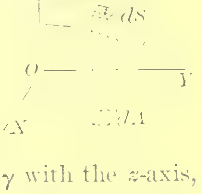
15. The same as Ex. 14 where three or four fractions are involved.

16. Note that the solution of $u_1 = \omega(x_1, y_1, z_1, u')$ for $u' = \omega_1(x_1, y_1, z_1, u_1)$ requires that $\partial\omega/\partial u'$ shall not vanish. Show that the hypothesis that J does not vanish in the region, is sufficient to show that at and in the neighborhood of each point (x, y, z, u) there must be at least one of the 16 derivatives of ϕ, ψ, χ, ω by x, y, z, u which does not vanish; and thus complete the proof of the text that in case $J \neq 0$ and the 16 derivatives exist and are continuous the change of variable is as given.

17. The intensity of light varies inversely as the square of the distance. Find the average intensity of illumination in a hemispherical dome lighted by a lamp at the top.

18. If the data be as in Ex. 12, p. 331, find the average density.

137. Surfaces and surface integrals. Consider a surface which has at each point a tangent plane that changes continuously from point to point of the surface. Consider also the projection of the surface upon a plane, say the xy-plane, and assume that a line perpendicular to the plane cuts the surface in only one point. Over any element dA of the projection there will be a small portion of the surface. If this small



portion were plane and if its normal made an angle γ with the z-axis, the area of the surface (p. 167) would be to its projection as 1 is to

$\cos\gamma$ and would be $\sec\gamma dA$. The value of $\cos\gamma$ may be read from (9) on page 96. This suggests that the quantity

$$S = \int \sec\gamma dA = \iint \left[1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2\right]^{\frac{1}{2}} dxdy \tag{15}$$

be taken as *the definition of the area of the surface*, where the double integral is extended over the projection of the surface; and this definition will be adopted. This definition is really dependent on the particular plane upon which the surface is projected; that the value of the area of the surface would turn out to be the same no matter what plane was used for projection is tolerably apparent, but will be proved later.

Let the area cut out of a hemisphere by a cylinder upon the radius of the hemisphere as diameter be evaluated. Here (or by geometry directly)

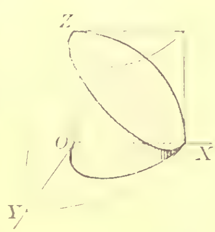
$$x^2 + y^2 + z^2 = a^2, \qquad \frac{\partial z}{\partial x} = -\frac{x}{z}, \qquad \frac{\partial z}{\partial y} = -\frac{y}{z},$$

$$S = \int \left[1 + \frac{x^2}{z^2} + \frac{y^2}{z^2}\right]^{\frac{1}{2}} dA = 2\int_{x=0}^{a} \int_{y=0}^{\sqrt{ax - x^2}} \frac{a}{\sqrt{a^2 - x^2 - y^2}} dydx.$$

This integral may be evaluated directly, but it is better to transform it to polar coördinates in the plane. Then

$$S = 2\int_{\phi=0}^{\frac{1}{2}\pi} \int_{r=0}^{a\cos\phi} \frac{a}{\sqrt{a^2 - r^2}} rdrd\phi = 2\int_0^{\frac{1}{2}\pi} a^2(1 - \sin\phi)d\phi = (\pi - 2)a^2.$$

It is clear that the half area which lies in the first octant could be projected upon the xz-plane and thus evaluated. The region over which the integration would extend is that between $x^2 + z^2 = a^2$ and the projection $z^2 + ax = a^2$ of the curve of intersection of the sphere and cylinder. The projection could also be made on the yz-plane. If the area of the cylinder between $z = 0$ and the sphere were desired, projection on $z = 0$ would be useless, projection on $x = 0$ would be involved owing to the overlapping of the projection on itself, but projection on $y = 0$ would be entirely feasible.



To show that the definition of area does not depend, except apparently, upon the plane of projection consider any second plane which makes an angle θ with the first. Let the line of intersection be the y-axis; then from a figure the new coördinate x' is

$$x' = x\cos\theta + z\sin\theta, \quad y = y, \quad \text{and} \quad J\left(\frac{x', y}{x, y}\right) = \frac{\partial x'}{\partial x} = \cos\theta + \frac{\partial z}{\partial x}\sin\theta,$$

$$S = \iint \frac{dxdy}{\cos\gamma} = \iint J\left(\frac{x, y}{x', y}\right)\frac{dx'dy}{\cos\gamma} = \iint \frac{dx'dy}{\cos\gamma(\cos\theta + p\sin\theta)}.$$

It remains to show that the denominator $\cos\gamma(\cos\theta + p\sin\theta) = \cos\gamma'$. Referred to the original axes the direction cosines of the normal are $-p : -q : 1$, and of

the z'-axis are $-\sin\theta : 0 : \cos\theta$. The cosine of the angle between these lines is therefore

$$\cos\gamma' = \frac{p\sin\theta + 0 + \cos\theta}{\sqrt{1 + p^2 + q^2}} = \frac{p\sin\theta + \cos\theta}{\sec\gamma} = \cos\gamma(\cos\theta + p\sin\theta).$$

Hence the new form of the area is the integral of $\sec\gamma' dA'$ and equals the old form.

The integrand $dS = \sec\gamma dA$ is called *the element of surface*. There are other forms such as $dS = \sec(r, n)r^2\sin\theta d\theta d\phi$, where (r, n) is the angle between the radius vector and the normal; but they are used comparatively little. The possession of an expression for the element of surface affords a means of computing *averages over surfaces*. For if $u = u(x, y, z)$ be any function of (x, y, z), and $z = f(x, y)$ any surface, the integral

$$\bar{u} = \frac{1}{S}\int u(x, y, z)dS = \frac{1}{S}\iint u(x, y, f)\sqrt{1 + p^2 + q^2}dxdy \qquad (16)$$

will be the average of u over the surface S. Thus the average height of a hemisphere is (for the surface average)

$$\bar{z} = \frac{1}{2\pi a^2}\int zdS = \frac{1}{2\pi a^2}\iint z\cdot\frac{a}{z}dxdy = \frac{1}{2\pi a^2}\cdot\pi a^3 = \frac{1}{2}a;$$

whereas the average height over the diametral plane would be $2/3$. This illustrates again the fact that the value of an average depends on the assumption made as to the weights.

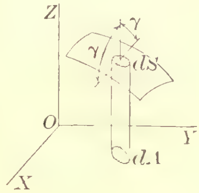
138. If a surface $z = f(x, y)$ be divided into elements ΔS_i, and the function $u(x, y, z)$ be formed for any point (ξ_i, η_i, ζ_i) of the element, and the sum $\Sigma u_i \Delta S_i$ be extended over all the elements, the limit of the sum as the elements become small in every direction is defined as the *surface integral* of the function over the surface and may be evaluated as

$$\lim\sum u(\xi_i, \eta_i, \zeta_i)\Delta S_i = \int u(x, y, z)dS$$
$$= \iint u[x, y, f(x, y)]\sqrt{1 + f_x'^2 + f_y'^2}dxdy. \qquad (17)$$

That the sum approaches a limit independently of how (ξ_i, η_i, ζ_i) is chosen in ΔS_i and how ΔS_i approaches zero follows from the fact that the element $u(\xi_i, \eta_i, \zeta_i)\Delta S_i$ of the sum differs uniformly from the integrand of the double integral by an infinitesimal of higher order, provided $u(x, y, z)$ be assumed continuous in (x, y, z) for points near the surface and $\sqrt{1 + f_x'^2 + f_y'^2}$ be continuous in (x, y) over the surface.

For many purposes it is more convenient to take as the normal form of the integrand of a surface integral, instead of udS, the

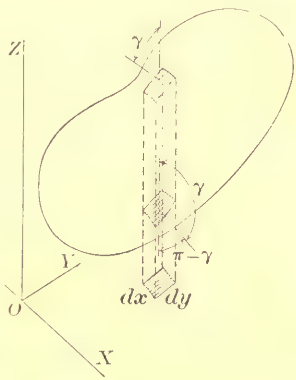
product $R \cos\gamma dS$ of a function $R(x, y, z)$ by the cosine of the inclination of the surface to the z-axis by the element dS of the surface. Then the integral may be evaluated *over either side* of the surface; for $R(x, y, z)$ has a definite value on the surface, dS is a positive quantity, but $\cos\gamma$ is positive or negative according as the normal is drawn on the upper or lower side of the surface. The value of the integral over the surface will be



$$\int R(x, y, z)\cos\gamma dS = \iint R dx dy \quad \text{or} \quad -\iint R dx dy \tag{18}$$

according as the evaluation is made over the upper or lower side. If the function $R(x, y, z)$ is continuous over the surface, these integrands will be finite even when the surface becomes perpendicular to the xy-plane, which might not be the case with an integrand of the form $u(x, y, z)dS$.

An integral of this sort may be evaluated over a closed surface. Let it be assumed that the surface is cut by a line parallel to the z-axis in a finite number of points, and for convenience let that number be two. Let the normal to the surface be taken constantly as the exterior normal (some take the interior normal with a resulting change of sign in some formulas), so that for the upper part of the surface $\cos\gamma > 0$ and for the lower part $\cos\gamma < 0$. Let $z = f_1(x, y)$ and $z = f_0(x, y)$ be the upper and lower values of z on the surface. Then the exterior integral over the closed surface will have the form



$$\int R\cos\gamma dS = \iint R[x, y, f_1(x, y)]dxdy - \iint R[x, y, f_0(x, y)]dxdy, \tag{18'}$$

where the double integrals are extended over the area of the projection of the surface on the xy-plane.

From this form of the surface integral over a closed surface it appears that a surface integral over a closed surface may be expressed as a volume integral over the volume inclosed by the surface.*

* Certain restrictions upon the functions and derivatives, as regards their becoming infinite and the like, must hold upon and within the surface. It will be quite sufficient if the functions and derivatives remain finite and continuous, but such extreme conditions are by no means necessary.

For by the rule for integration,

$$\iiint_{z=f_0(x,y)}^{z=f_1(x,y)} \frac{\partial R}{\partial z}\,dzdxdy = \iint R(x, y, z)\Big|_{z=f_0(x,y)}^{z=f_1(x,y)} dxdy.$$

Hence
$$\int_{\bigcirc} R\cos\gamma dS = \int \frac{\partial R}{\partial z}\,dV$$

or
$$\iint_{\bigcirc} Rdxdy = \iiint \frac{\partial R}{\partial z}\,dxdydz \tag{19}$$

if the symbol ○ be used to designate a closed surface, and if the double integral on the left of (19) be understood to stand for either side of the equality (18′). In a similar manner

$$\int P\cos\alpha dS = \iint_{\bigcirc} Pdydz = \iiint \frac{\partial P}{\partial x}\,dxdydz = \int \frac{\partial P}{\partial x}\,dV,$$

$$\int Q\cos\beta dS = \iint_{\bigcirc} Qdxdz = \iiint \frac{\partial Q}{\partial y}\,dydxdz = \int \frac{\partial Q}{\partial y}\,dV. \tag{19′}$$

Then
$$\int_{\bigcirc} (P\cos\alpha + Q\cos\beta + R\cos\gamma)\,dS = \int \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dV$$

or
$$\iint_{\bigcirc} (Pdydz + Qdzdx + Rdxdy) = \iiint \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dxdydz \tag{20}$$

follows immediately by merely adding the three equalities. Any one of these equalities (19), (20) is sometimes called *Gauss's Formula*, sometimes *Green's Lemma*, sometimes *the divergence formula* owing to the interpretation below.

The interpretation of Gauss's Formula (20) by vectors is important. From the viewpoint of vectors the element of surface is a vector $d\mathbf{S}$ directed along the exterior normal to the surface (§ 76). Construct the vector function

$$\mathbf{F}(x, y, z) = \mathbf{i}P(x, y, z) + \mathbf{j}Q(x, y, z) + \mathbf{k}R(x, y, z).$$

Let $\quad d\mathbf{S} = (\mathbf{i}\cos\alpha + \mathbf{j}\cos\beta + \mathbf{k}\cos\gamma)\,dS = \mathbf{i}dS_x + \mathbf{j}dS_y + \mathbf{k}dS_z,$

where dS_x, dS_y, dS_z are the projections of dS on the coördinate planes

Then
$$P\cos\alpha dS + Q\cos\beta dS + R\cos\gamma dS = \mathbf{F}\cdot d\mathbf{S}$$

and
$$\iint (Pdydz + Qdxdz + Rdxdy) = \int \mathbf{F}\cdot d\mathbf{S},$$

where dS_x, dS_y, dS_z have been replaced by the elements $dydz$, $dxdz$, $dxdy$, which would be used to evaluate the integrals in rectangular coördinates,

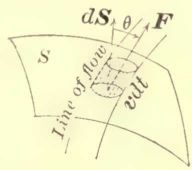
without at all implying that the projections dS_x, dS_y, dS_z are actually rectangular. The combination of partial derivatives

$$\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}=\operatorname{div}\mathbf{F}=\nabla\cdot\mathbf{F}, \tag{21}$$

where $\nabla\cdot\mathbf{F}$ is the symbolic scalar product of ∇ and $\mathbf{F}$ (Ex. 9 below), is called the *divergence* of $\mathbf{F}$. Hence (20) becomes

$$\int \operatorname{div}\mathbf{F}dV=\int\nabla\cdot\mathbf{F}dV=\int\mathbf{F}\cdot d\mathbf{S}. \tag{20'}$$

Now the function $\mathbf{F}(x, y, z)$ is such that at each point (x, y, z) of space a vector is defined. Such a function is seen in the velocity in a moving fluid such as air or water. The picture of a scalar function $u(x, y, z)$ was by means of the surfaces $u=$ const.; the picture of a vector function $\mathbf{F}(x, y, z)$ may be found in the system of curves tangent to the vector, the stream lines in the fluid if $\mathbf{F}$ be the velocity. For the immediate purposes it is better to consider the function $\mathbf{F}(x, y, z)$ as the flux $D\mathbf{v}$, the product of the density in the fluid by the velocity. With this interpretation the rate at which the fluid flows through an element of surface $d\mathbf{S}$ is $D\mathbf{v}\cdot d\mathbf{S}=\mathbf{F}\cdot d\mathbf{S}$. For in the time dt the fluid will advance along a stream line by the amount $\mathbf{v}dt$ and the volume of the cylindrical volume of fluid which advances through the surface will be $\mathbf{v}\cdot d\mathbf{S}dt$. Hence $\Sigma D\mathbf{v}\cdot d\mathbf{S}$ will be the rate of diminution of the amount of fluid within the closed surface.



As the amount of fluid in an element of volume dV is DdV, the rate of diminution of the fluid in the element of volume is $-\partial D/\partial t$ where $\partial D/\partial t$ is the rate of increase of the density D at a point within the element. The total rate of diminution of the amount of fluid within the whole volume is therefore $-\Sigma\partial D/\partial t\,dV$. Hence, by virtue of the principle of the indestructibility of matter,

$$\int\mathbf{F}\cdot d\mathbf{S}=\int D\mathbf{v}\cdot d\mathbf{S}=-\int\frac{\partial D}{\partial t}dV. \tag{20''}$$

Now if v_x, v_y, v_z be the components of $\mathbf{v}$ so that $P=Dv_x$, $Q=Dv_y$, $R=Dv_z$ are the components of $\mathbf{F}$, a comparison of (21), (20′), (20″) shows that the integrals of $-\partial D/\partial t$ and div $\mathbf{F}$ are always equal, and hence the integrands,

$$-\frac{\partial D}{\partial t}\quad\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}=\frac{\partial Dv_x}{\partial x}+\frac{\partial Dv_y}{\partial y}+\frac{\partial Dv_z}{\partial z},$$

are equal; that is, the sum $P'_x+Q'_y+R'_z$ represents the rate of diminution of density when $\mathbf{i}P+\mathbf{j}Q+\mathbf{k}R$ is the flux vector; this combination is called the divergence of the vector, no matter what the vector $\mathbf{F}$ really represents.

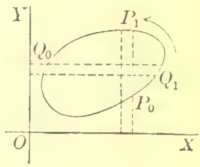
139. Not only may a surface integral be stepped up to a volume integral, but a line integral around a closed curve may be stepped up into a surface integral over a surface which spans the curve. To begin

with the simple case of a line integral in a plane, note that by the same reasoning as above

$$\int_{\bigcirc} P dx = \iint -\frac{\partial P}{\partial y}\, dx dy, \qquad \int_{\bigcirc} Q dy = \iint \frac{\partial Q}{\partial x}\, dx dy,$$

$$\int_{\bigcirc} [P(x, y)\, dx + Q(x, y)\, dy] = \iint \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx dy. \tag{22}$$

This is sometimes called *Green's Lemma for the plane* in distinction to the general Green's Lemma for space. The opposite signs must be taken to preserve the direction of the line integral about the contour. This result may be used to establish the rule for transforming a double integral by the change of variable $x = \phi(u, v)$, $y = \psi(u, v)$. For



$$A = \int_{\bigcirc} x dy = \pm \int_{\bigcirc} x \frac{\partial y}{\partial u}\, du + x \frac{\partial y}{\partial v}\, dv$$

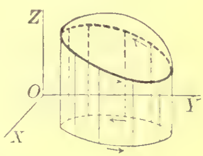
$$= \pm \iint \left[\frac{\partial}{\partial u}\left(x \frac{\partial y}{\partial v}\right) - \frac{\partial}{\partial v}\left(x \frac{\partial y}{\partial u}\right)\right] du dv$$

$$= \pm \iint \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right) du dv$$

$$= \pm \iint J\left(\frac{x, y}{u, v}\right) du dv = \iint |J|\, du dv.$$

(The double signs have to be introduced at first to allow for the case where J is negative.) The element of area $dA = |J|\, dudv$ is therefore established.

To obtain the formula for the conversion of a line integral in space to a surface integral, let $P(x, y, z)$ be given and let $z = f(x, y)$ be a surface spanning the closed curve $\bigcirc$. Then by virtue of $z = f(x, y)$, the function $P(x, y, z) = P_1(x, y)$ and



$$\int_{\bigcirc} P dx = \int_{\bigcirc'} P_1 dx = \iint -\frac{\partial P_1}{\partial y}\, dx dy = -\iint \left(\frac{\partial P}{\partial y} + \frac{\partial z}{\partial y}\frac{\partial P}{\partial z}\right) dx dy,$$

where $\bigcirc'$ denotes the projection of $\bigcirc$ on the xy-plane. Now the final double integral may be transformed by the introduction of the cosines of the normal direction to $z = f(x, y)$.

$$\cos\beta : \cos\gamma = -q : 1, \quad dxdy = \cos\gamma dS, \quad q dxdy = -\cos\beta dS = -dxdz.$$

Then $-\iint\left(\frac{\partial P}{\partial y}+q\frac{\partial P}{\partial z}\right)dxdy=\iint\left(\frac{\partial P}{\partial z}dxdz-\frac{\partial P}{\partial y}dxdy\right)=\int_{\circlearrowleft}Pdx.$

If this result and those obtained by permuting the letters be added,

$$\int_{\circlearrowleft}(Pdx+Qdy+Rdz)$$
$$=\iint\left[\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)dydz+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)dxdz+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dxdy\right]. \quad (23)$$

This is known as *Stokes's Formula* and is of especial importance in hydromechanics and the theory of electromagnetism. Note that the line integral is carried around the rim of the surface in the direction which appears positive to one standing upon that side of the surface over which the surface integral is extended.

Again the vector interpretation of the result is valuable. Let

$$\mathbf{F}(x,y,z)=\mathbf{i}P(x,y,z)+\mathbf{j}Q(x,y,z)+\mathbf{k}R(x,y,z),$$

$$\text{curl }\mathbf{F}=\mathbf{i}\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)+\mathbf{j}\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)+\mathbf{k}\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right). \quad (24)$$

Then $$\int_{\circlearrowleft}\mathbf{F}\cdot d\mathbf{r}=\int\text{curl }\mathbf{F}\cdot d\mathbf{S}=\int\nabla\times\mathbf{F}\cdot d\mathbf{S}, \quad (23')$$

where $\nabla\times\mathbf{F}$ is the symbolic vector product of ∇ and $\mathbf{F}$ (Ex. 9, below), is the form of Stokes's Formula; that is, the line integral of a vector around a closed curve is equal to the surface integral of the curl of the vector, as defined by (24), around any surface which spans the curve. If the line integral is zero about every closed curve, the surface integral must vanish over every surface. It follows that curl $\mathbf{F}=0$. For if the vector curl $\mathbf{F}$ failed to vanish at any point, a small plane surface $d\mathbf{S}$ perpendicular to the vector might be taken at that point and the integral over the surface would be approximately $|\text{curl }\mathbf{F}|dS$ and would fail to vanish, — thus contradicting the hypothesis. Now the vanishing of the vector curl $\mathbf{F}$ requires the vanishing

$$R'_y-Q'_z=0,\qquad P'_z-R'_x=0,\qquad Q'_x-P'_y=0$$

of each of its components. Thus may be derived the condition that $Pdx+Qdy+Rdz$ be an exact differential.

If $\mathbf{F}$ be interpreted as the velocity $\mathbf{v}$ in a fluid, the integral

$$\int\mathbf{v}\cdot d\mathbf{r}=\int v_xdx+v_ydy+v_zdz$$

of the component of the velocity along a curve, whether open or closed, is called the *circulation* of the fluid along the curve; it might be more natural to define

the integral of the flux $D\mathbf{v}$ along the curve as the circulation, but this is not the convention. Now if the velocity be that due to rotation with the angular velocity $\mathbf{a}$ about a line through the origin, the circulation in a closed curve is readily computed. For

$$\mathbf{v} = \mathbf{a}\times\mathbf{r}. \qquad \int_{\circ} \mathbf{v}\cdot d\mathbf{r} = \int_{\circ} \mathbf{a}\times\mathbf{r}\cdot d\mathbf{r} = \int_{\circ} \mathbf{a}\cdot\mathbf{r}\times d\mathbf{r} = \mathbf{a}\cdot\int_{\circ} \mathbf{r}\times d\mathbf{r} = 2\,\mathbf{a}\cdot\mathbf{A}.$$

The circulation is therefore the product of twice the angular velocity and the area of the surface inclosed by the curve. If the circuit be taken indefinitely small, the integral is $2\,\mathbf{a}\cdot d\mathbf{S}$ and a comparison with (23′) shows that curl $\mathbf{v} = 2\,\mathbf{a}$; that is, the curl of the velocity due to rotation about an axis is twice the angular velocity and is constant in magnitude and direction all over space. The general motion of a fluid is not one of uniform rotation about any axis; in fact if a small element of fluid be considered and an interval of time δt be allowed to elapse, the element will have moved into a new position, will have been somewhat deformed owing to the motion of the fluid, and will have been somewhat rotated. The vector curl $\mathbf{v}$, as defined in (24), may be shown to give twice the instantaneous angular velocity of the element at each point of space.

EXERCISES

1. Find the areas of the following surfaces:

(α) cylinder $x^2 + y^2 - ax = 0$ included by the sphere $x^2 + y^2 + z^2 = a^2$.
(β) $x/a + y/b + z/c = 1$ in first octant. (γ) $x^2 + y^2 + z^2 = a^2$ above $r = a\cos n\phi$.
(δ) sphere $x^2 + y^2 + z^2 = a^2$ above a square $|x| \leqq b$, $|y| \leqq b$, $b < \frac{1}{2}\sqrt{2}\,a$.
(ϵ) $z = xy$ over $x^2 + y^2 = a^2$. (ζ) $2\,az = x^2 - y^2$ over $r^2 = a^2\cos\phi$.
(η) $z^2 + (x\cos\alpha + y\sin\alpha)^2 = a^2$ in first octant. (θ) $z = xy$ over $r^2 = \cos 2\,\phi$.
(ι) cylinder $x^2 + y^2 = a^2$ intercepted by equal cylinder $y^2 + z^2 = a^2$.

2. Compute the following superficial averages:

(α) latitude of places north of the equator, *Ans.* $32\frac{7}{10}°$.
(β) ordinate in a right circular cone $h^2(x^2 + y^2) - a^2(z - h)^2 = 0$.
(γ) illumination of a hollow spherical surface by a light at a point of it,
(δ) illumination of a hemispherical surface by a distant light,
(ϵ) rectilinear distance of points north of equator from north pole.

3. A theorem of Pappus: If a closed or open plane curve be revolved about an axis in its plane, the area of the surface generated is equal to the product of the length of the curve by the distance described by the center of gravity of the curve. The curve shall not cut the axis. Prove either analytically or by infinitesimal analysis. Apply to various figures in which two of the three quantities, length of curve, area of surface, position of center of gravity, are known, to find the third. Compare Ex. 27, p. 332.

4. The surface integrals are to be evaluated over the closed surfaces by expressing them as volume integrals. Try also direct calculation:

(α) $\iint (x^2dydz + xydxdy + xzdxdz)$ over the spherical surface $x^2 + y^2 + z^2 = a^2$.

(β) $\iint (x^2dydz + y^2dxdz + z^2dxdy)$, cylindrical surface $x^2 + y^2 = a^2$, $z = \pm h$.

(γ) $\iint[(x^2-yz)dydz-2xydxdz+dxdy]$ over the cube $0 \leqq x, y, z \leqq a$,

(δ) $\iint xdydz=\iint ydxdz=\iint zdxdy=\frac{1}{3}\iint(xdydz+ydxdz+zdxdy)=V$,

(ϵ) Calculate the line integrals of Ex. 8, p. 297, around a closed path formed by two paths there given, by applying Green's Lemma (22) and evaluating the resulting double integrals.

5. If $x=\phi_1(u, v)$, $y=\phi_2(u, v)$, $z=\phi_3(u, v)$ are the parametric equations of a surface, the direction ratios of the normal are (see Ex. 15, p. 135)

$$\cos\alpha:\cos\beta:\cos\gamma=J_1:J_2:J_3 \quad \text{if} \quad J_i=J\left(\frac{\phi_{i+1}, \phi_{i+2}}{u, v}\right).$$

Show 1° that the area of a surface may be written as

$$S=\iint\frac{\sqrt{J_1^2+J_2^2+J_3^2}}{|J_3|}dxdy=\iint\sqrt{J_1^2+J_2^2+J_3^2}\,dudv=\iint\sqrt{EG-F^2}dudv,$$

where $\quad E=\sum\left(\frac{\partial\phi_i}{\partial u}\right)^2, \quad G=\sum\left(\frac{\partial\phi_i}{\partial v}\right)^2, \quad F=\sum\frac{\partial\phi_i}{\partial u}\frac{\partial\phi_i}{\partial v},$

and $\quad ds^2=Edu^2+2\,Fdudv+Gdv^2.$

Show 2° that the surface integral of the first type becomes merely

$$\iint f(x, y, z)\sec\gamma dxdy=\iint f(\phi_1, \phi_2, \phi_3)\sqrt{EG-F^2}dudv,$$

and determine the integrand in the case of the developable surface of Ex. 17, p. 143.

Show 3° that if $x=f_1(\xi, \eta, \zeta)$, $y=f_2(\xi, \eta, \zeta)$, $z=f_3(\xi, \eta, \zeta)$ is a transformation of space which transforms the above surface into a new surface $\xi=\psi_1(u, v)$, $\eta=\psi_2(u, v)$, $\zeta=\psi_3(u, v)$, then

$$J\left(\frac{x, y}{u, v}\right)=J\left(\frac{x, y}{\xi, \eta}\right)J\left(\frac{\xi, \eta}{u, v}\right)+J\left(\frac{x, y}{\eta, \zeta}\right)J\left(\frac{\eta, \zeta}{u, v}\right)+J\left(\frac{x, y}{\zeta, \xi}\right)J\left(\frac{\zeta, \xi}{u, v}\right).$$

Show 4° that the surface integral of the second type becomes

$$\iint Rdxdy=\iint RJ\left(\frac{x, y}{u, v}\right)dudv$$
$$=\iint R\left[J\left(\frac{x, y}{\eta, \zeta}\right)d\eta d\zeta+J\left(\frac{x, y}{\zeta, \xi}\right)d\zeta d\xi+J\left(\frac{x, y}{\xi, \eta}\right)d\xi d\eta\right],$$

where the integration is now in terms of the new variables ξ, η, ζ in place of x, y, z.

Show 5° that when $R=z$ the double integral above may be transformed by Green's Lemma in such a manner as to establish the formula for change of variables in triple integrals.

6. Show that for vector surface integrals $\int_{\bigcirc} Ud\mathbf{S}=\int\nabla UdV$.

7. *Solid angle as a surface integral.* The area cut out from the unit sphere by a cone with its vertex at the center of the sphere is called the *solid angle* ω subtended at the vertex of the cone. The solid angle may also be defined as the ratio of the area cut out upon any sphere concentric with the vertex of the cone, to the square of the radius of the sphere (compare the definition of the angle between two lines

in radians). Show geometrically (compare Ex. 16, p. 297) that the infinitesimal solid angle $d\omega$ of the cone which joins the origin $r = 0$ to the periphery of the element dS of a surface is $d\omega = \cos(r, n)\, dS/r^2$, where (r, n) is the angle between the radius produced and the outward normal to the surface. Hence show

$$\omega = \int \frac{\cos(r, n)}{r^2} dS = \int \frac{\mathbf{r}\cdot d\mathbf{S}}{r^3} = \int \frac{1}{r^2}\frac{dr}{dn} dS = -\int \frac{d}{dn}\frac{1}{r} dS = -\int d\mathbf{S}\cdot\nabla \frac{1}{r},$$

where the integrals extend over a surface, is the solid angle subtended at the origin by that surface. Infer further that

$$-\oint \frac{d}{dn}\frac{1}{r} dS = 4\pi \quad \text{or} \quad -\oint \frac{d}{dn}\frac{1}{r} dS = 0 \quad \text{or} \quad -\oint \frac{d}{dn}\frac{1}{r} dS = \theta$$

according as the point $r = 0$ is within the closed surface or outside it or upon it at a point where the tangent planes envelop a cone of solid angle θ (usually 2π). Note that the formula may be applied at any point (ξ, η, ζ) if

$$r^2 = (\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2$$

where (x, y, z) is a point of the surface.

8. *Gauss's Integral.* Suppose that at $\mathbf{r} = 0$ there is a particle of mass m which *attracts* according to the Newtonian Law $F = m/r^2$. Show that the potential is $V = -m/r$ so that $\mathbf{F} = -\nabla V$. The induction or flux (see Ex. 19, p. 308) of the force $\mathbf{F}$ outward across the element $d\mathbf{S}$ of a surface is by definition $-F\cos(F, n)\, dS = \mathbf{F}\cdot d\mathbf{S}$. Show that the total induction or flux of $\mathbf{F}$ across a surface is the surface integral

$$\int \mathbf{F}\cdot d\mathbf{S} = -\int d\mathbf{S}\cdot\nabla V = -\int \frac{dV}{dn} dS = m\int d\mathbf{S}\cdot\nabla \frac{1}{r};$$

and
$$m = \frac{-1}{4\pi}\oint \mathbf{F}\cdot d\mathbf{S} = \frac{1}{4\pi}\oint d\mathbf{S}\cdot\nabla V = \frac{-1}{4\pi}\oint \frac{d}{dn}\frac{m}{r} dS,$$

where the surface integral extends over a surface surrounding a point $\mathbf{r} = 0$, is the formula for obtaining the mass m within the surface from the field of force $\mathbf{F}$ which is set up by the mass. If there are several masses $m_1, m_2, \cdots$ situated at points (ξ_1, η_1, ζ_1), (ξ_2, η_2, ζ_2), $\cdots$, let

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \cdots, \qquad V = V_1 + V_2 + \cdots,$$
$$V_i = -m\left[(\xi_i - x_i)^2 + (\eta_i - y_i)^2 + (\zeta_i - z_i)^2\right]^{-\frac{1}{2}}$$

be the force and potential at (x, y, z) due to the masses. Show that

$$\frac{-1}{4\pi}\oint \mathbf{F}\cdot d\mathbf{S} = \frac{1}{4\pi}\oint d\mathbf{S}\cdot\nabla V = -\frac{1}{4\pi}\sum \oint \frac{d}{dn}\frac{1}{r_i} dS = \sum{}' m_i = M, \qquad (25)$$

where Σ extends over all the masses and Σ' over all the masses within the surface (none being on it), gives the total mass M within the surface. The integral (25) which gives the mass within a surface as a surface integral is known as Gauss's Integral. If the force were repulsive (as in electricity and magnetism) instead of attracting (as in gravitation), the results would be $V = m/r$ and

$$\frac{1}{4\pi}\oint \mathbf{F}\cdot d\mathbf{S} = \frac{-1}{4\pi}\oint d\mathbf{S}\cdot\nabla V = \frac{-1}{4\pi}\sum \oint \frac{d}{dn}\frac{m_i}{r_i} dS = \sum{}' m_i = M. \qquad (25')$$

9. If $\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$ be the operator defined on page 172, show

$$\nabla\cdot\mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}, \qquad \nabla\times\mathbf{F} = \mathbf{i}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + \mathbf{j}\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + \mathbf{k}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)$$

by formal operation on $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$. Show further that

$$\nabla\times\nabla U = 0, \qquad \nabla\cdot\nabla\times\mathbf{F} = 0, \qquad (\nabla\cdot\nabla)(*) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)(*),$$

$$\nabla\times(\nabla\times\mathbf{F}) = \nabla(\nabla\cdot\mathbf{F}) - (\nabla\cdot\nabla)\mathbf{F} \qquad \text{(write the Cartesian form).}$$

Show that $(\nabla\cdot\nabla)\,U = \nabla\cdot(\nabla U)$. If $\mathbf{u}$ is a constant unit vector, show

$$(\mathbf{u}\cdot\nabla)\,\mathbf{F} = \frac{\partial\mathbf{F}}{\partial x}\cos\alpha + \frac{\partial\mathbf{F}}{\partial y}\cos\beta + \frac{\partial\mathbf{F}}{\partial z}\cos\gamma = \frac{d\mathbf{F}}{ds}$$

is the directional derivative of $\mathbf{F}$ in the direction $\mathbf{u}$. Show $(d\mathbf{r}\cdot\nabla)\,\mathbf{F} = d\mathbf{F}$.

10. *Green's Formula* (space). Let $F(x, y, z)$ and $G(x, y, z)$ be two functions so that ∇F and ∇G become two vector functions and $F\nabla G$ and $G\nabla F$ two other vector functions. Show

$$\nabla\cdot(F\nabla G) = \nabla F\cdot\nabla G + F\nabla\cdot\nabla G, \qquad \nabla\cdot(G\nabla F) = \nabla F\cdot\nabla G + G\nabla\cdot\nabla F,$$

or
$$\frac{\partial}{\partial x}\left(F\frac{\partial G}{\partial x}\right) + \frac{\partial}{\partial y}\left(F\frac{\partial G}{\partial y}\right) + \frac{\partial}{\partial z}\left(F\frac{\partial G}{\partial z}\right)$$
$$= \frac{\partial F}{\partial x}\frac{\partial G}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial G}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial G}{\partial z} + F\left(\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} + \frac{\partial^2 G}{\partial z^2}\right),$$

and the similar expressions which are the Cartesian equivalents of the above vector forms. Apply Green's Lemma or Gauss's Formula to show

$$\int_{\bigcirc} F\nabla G\cdot d\mathbf{S} = \int \nabla F\cdot\nabla G\, dV + \int F\nabla\cdot\nabla G\, dV, \tag{26}$$

$$\int_{\bigcirc} G\nabla F\cdot d\mathbf{S} = \int \nabla F\cdot\nabla G\, dV + \int G\nabla\cdot\nabla F\, dV, \tag{26'}$$

$$\int_{\bigcirc} (F\nabla G - G\nabla F)\cdot d\mathbf{S} = \int (F\nabla\cdot\nabla G - G\nabla\cdot\nabla F)\, dV, \tag{26''}$$

or $$\int_{\bigcirc} F\frac{dG}{dn}dS = \int\left(\frac{\partial F}{\partial x}\frac{\partial G}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial G}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial G}{\partial z}\right)dV + \int F\left(\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} + \frac{\partial^2 G}{\partial z^2}\right)dV,$$

$$\int_{\bigcirc}\left(F\frac{dG}{dn} - G\frac{dF}{dn}\right)dS = \int\left[F\left(\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} + \frac{\partial^2 G}{\partial z^2}\right) - G\left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2}\right)\right]dV.$$

The formulas (26), (26′), (26″) are known as *Green's Formulas;* in particular the first two are asymmetric and the third symmetric. The ordinary Cartesian forms of (26) and (26″) are given. The expression $\partial^2F/\partial x^2 + \partial^2F/\partial y^2 + \partial^2F/\partial z^2$ is often written as ΔF for brevity ; the vector form is $\nabla\cdot\nabla F$.

11. From the fact that the integral of $\mathbf{F}\cdot d\mathbf{r}$ has opposite values when the curve is traced in opposite directions, show that the integral of $\nabla\times\mathbf{F}$ over a closed surface vanishes and that the integral of $\nabla\cdot\nabla\times\mathbf{F}$ over a volume vanishes. Infer that $\nabla\cdot\nabla\times\mathbf{F} = 0$.

12. Reduce the integral of $\nabla\times\nabla U$ over any (open) surface to the difference in the values of U at two same points of the bounding curve. Hence infer $\nabla\times\nabla U = 0$.

13. Comment on the remark that the line integral of a vector, integral of $\mathbf{F}\cdot d\mathbf{r}$, is around a curve and *along* it, whereas the surface integral of a vector, integral of $\mathbf{F}\cdot d\mathbf{S}$, is over a surface but *through* it. Compare Ex. 7 with Ex. 16 of p. 297. In particular give vector forms of the integrals in Ex. 16, p. 297, analogous to those of Ex. 7 by using as the element of the curve a normal $d\mathbf{n}$ equal in length to $d\mathbf{r}$, instead of $d\mathbf{r}$.

14. If in $\mathbf{F} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, the functions P, Q depend only on x, y and the function $R = 0$, apply Gauss's Formula to a cylinder of unit height upon the xy-plane to show that

$$\int \nabla\cdot\mathbf{F}\,dV = \int \mathbf{F}\cdot d\mathbf{S} \quad \text{becomes} \quad \int\!\!\int \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\,dy = \int \mathbf{F}\cdot d\mathbf{n},$$

where $d\mathbf{n}$ has the meaning given in Ex. 13. Show that numerically $\mathbf{F}\cdot d\mathbf{n}$ and $\mathbf{F}\times d\mathbf{r}$ are equal, and thus obtain Green's Lemma for the plane (22) as a special case of (20). Derive Green's Formula (Ex. 10) for the plane.

15. If $\int \mathbf{F}\cdot d\mathbf{r} = \int \mathbf{G}\cdot d\mathbf{S}$, show that $\int (\mathbf{G} - \nabla\times\mathbf{F})\cdot d\mathbf{S} = 0$. Hence infer that if these relations hold for every surface and its bounding curve, then $\mathbf{G} = \nabla\times\mathbf{F}$. Ampère's Law states that the integral of the magnetic force $\mathbf{H}$ about any circuit is equal to 4π times the flux of the electric current $\mathbf{C}$ through the circuit, that is, through any surface spanning the circuit. Faraday's Law states that the integral of the electromotive force $\mathbf{E}$ around any circuit is the negative of the time rate of flux of the magnetic induction $\mathbf{B}$ through the circuit. Phrase these laws as integrals and convert into the form

$$4\pi\mathbf{C} = \operatorname{curl}\mathbf{H}, \qquad -\dot{\mathbf{B}} = \operatorname{curl}\mathbf{E}.$$

16. By formal expansion prove $\nabla\cdot(\mathbf{E}\times\mathbf{H}) = \mathbf{H}\cdot\nabla\times\mathbf{E} - \mathbf{E}\cdot\nabla\times\mathbf{H}$. Assume $\nabla\times\mathbf{E} = -\dot{\mathbf{H}}$ and $\nabla\times\mathbf{H} = \dot{\mathbf{E}}$ and establish Poynting's Theorem that

$$\int (\mathbf{E}\times\mathbf{H})\cdot d\mathbf{S} = -\frac{\partial}{\partial t}\int \frac{1}{2}(\mathbf{E}\cdot\mathbf{E} + \mathbf{H}\cdot\mathbf{H})\,dV.$$

17. The "equation of continuity" for fluid motion is

$$\frac{\partial D}{\partial t} + \frac{\partial Dv_x}{\partial x} + \frac{\partial Dv_y}{\partial y} + \frac{\partial Dv_z}{\partial z} = 0 \quad \text{or} \quad \frac{dD}{dt} + D\left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}\right) = 0,$$

where D is the density, $\mathbf{v} = \mathbf{i}v_x + \mathbf{j}v_y + \mathbf{k}v_z$ is the velocity, $\partial D/\partial t$ is the rate of change of the density at a point, and dD/dt is the rate of change of density as one moves with the fluid (Ex. 14, p. 101). Explain the meaning of the equation in view of the work of the text. Show that for fluids of constant density $\nabla\cdot\mathbf{v} = 0$.

18. If $\mathbf{f}$ denotes the acceleration of the particles of a fluid, and if $\mathbf{F}$ is the external force acting per unit mass upon the elements of fluid, and if p denotes the pressure in the fluid, show that the equation of motion for the fluid within any surface may be written as

$$\sum \mathbf{f}D\,dV = \sum \mathbf{F}D\,dV - \sum p\,d\mathbf{S} \quad \text{or} \quad \int \mathbf{f}D\,dV = \int \mathbf{F}D\,dV - \int p\,d\mathbf{S}.$$

where the summations or integrations extend over the volume or its bounding surface and the pressures (except those acting on the bounding surface inward) may be disregarded. (See the first half of § 80.)

19. By the aid of Ex. 6 transform the surface integral in Ex. 18 and find

$$\int D\mathbf{f}dV = \int (D\mathbf{F} - \nabla p)\,dV \quad \text{or} \quad \frac{d^2\mathbf{r}}{dt^2} = \mathbf{F} - \frac{1}{D}\nabla p$$

as the equations of motion for a fluid, where $\mathbf{r}$ is the vector to any particle. Prove

$$(\alpha)\ \frac{d^2\mathbf{r}}{dt^2} = \frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\,\mathbf{v} = \frac{\partial \mathbf{v}}{\partial t} - \mathbf{v}\times\nabla\times\mathbf{v} + \frac{1}{2}\nabla(\mathbf{v}\cdot\mathbf{v}),$$

$$(\beta)\ \frac{d}{dt}(d\mathbf{r}\cdot\mathbf{v}) = d\mathbf{r}\cdot\frac{d\mathbf{v}}{dt} + d\frac{d\mathbf{r}}{dt}\cdot\mathbf{v} = d\mathbf{r}\cdot\frac{d^2\mathbf{r}}{dt^2} + \frac{1}{2}d(\mathbf{v}\cdot\mathbf{v}).$$

20. If $\mathbf{F}$ is derivable from a potential, so that $\mathbf{F} = -\nabla U$, and if the density is a function of the pressure, so that $dp/D = dP$, show that the equations of motion are

$$\frac{\partial \mathbf{v}}{\partial t} - \mathbf{v}\times\nabla\times\mathbf{v} = -\nabla\left(U + P + \frac{1}{2}v^2\right), \quad \text{or} \quad \frac{d}{dt}(\mathbf{v}\cdot d\mathbf{r}) = -d\left(U + P - \frac{1}{2}v^2\right)$$

after multiplication by $d\mathbf{r}$. The first form is Helmholtz's, the second is Kelvin's. Show

$$\int_{a,\,b,\,c}^{x,\,y,\,z} \frac{d}{dt}(\mathbf{v}\cdot d\mathbf{r}) = \frac{d}{dt}\int_{a,\,b,\,c}^{x,\,y,\,z} \mathbf{v}\cdot d\mathbf{r} = -\left[U + P - \frac{1}{2}v^2\right]_{a,\,b,\,c}^{x,\,y,\,z} \text{ and } \int_{\bigcirc} \mathbf{v}\cdot d\mathbf{r} = \text{const.}$$

In particular explain that as the differentiation d/dt follows the particles in their motion (in contrast to $\partial/\partial t$, which is executed at a single point of space), the integral must do so if the order of differentiation and integration is to be interchangeable. Interpret the final equation as stating that the circulation in a curve which moves with the fluid is constant.

21. If $\dfrac{\partial^2 U}{\partial x^2} + \dfrac{\partial^2 U}{\partial y^2} + \dfrac{\partial^2 U}{\partial z^2} = 0$, show $\displaystyle\int\left[\left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial z}\right)^2\right]dV = \int_{\bigcirc} U\frac{dU}{dn}dS.$

22. Show that, apart from the proper restrictions as to continuity and differentiability, the necessary and sufficient condition that the surface integral

$$\iint Pdydz + Qdzdx + Rdxdy = \int_{\bigcirc} pdx + qdy + rdz$$

depends only on the curve bounding the surface is that $P'_x + Q'_y + R'_z = 0$. Show further that in this case the surface integral reduces to the line integral given above, provided p, q, r are such functions that $r'_y - q'_z = P$, $p'_z - r'_x = Q$, $q'_x - p'_y = R$. Show finally that these differential equations for p, q, r may be satisfied by

$$p = \int_{z_0}^{z} Qdz - \int R(x,\ y,\ z_0)\,dy, \qquad q = -\int_{z_0}^{z} Pdz, \qquad r = 0;$$

and determine by inspection alternative values of p, q, r.

CHAPTER XIII

ON INFINITE INTEGRALS

140. Convergence and divergence. The definite integral, and hence for theoretical purposes the indefinite integral, has been defined,

$$\int_a^b f(x)\,dx, \qquad F(x) = \int_a^x f(x)\,dx,$$

when the function $f(x)$ is *limited* in the interval a to b, or a to x; the proofs of various propositions have depended essentially on the fact that *the integrand remained finite over the finite interval of integration* (§§ 16–17, 28–30). Nevertheless problems which call for the determination of the area between a curve and its asymptote, say the area under the witch or cissoid,

$$\int_{-\infty}^{+\infty} \frac{8a^3dx}{x^2+4a^2} = 4a^2\tan^{-1}\frac{x}{2a}\Bigg|_{-\infty}^{+\infty} = 4\pi a^2, \qquad 2\int_0^{2a}\frac{x^{\frac{3}{2}}dx}{\sqrt{2a-x}} = 3\pi a^2,$$

have arisen and have been treated as a matter of course.* The integrals of this sort require some special attention.

When the integrand of a definite integral becomes infinite within or at the extremities of the interval of integration, or when one or both of the limits of integration become infinite, the integral is called an infinite integral and is defined, not as the limit of a sum, but as the limit of an integral with a variable limit, that is, as the limit of a function. Thus

$$\int_a^\infty f(x)\,dx = \lim_{x=\infty}\left[F(x) = \int_a^x f(x)dx\right], \qquad \text{infinite upper limit,}$$

$$\int_a^b f(x)\,dx = \lim_{x\doteq b}\left[F(x) = \int_a^x f(x)\,dx\right], \qquad \text{integrand } f(b)=\infty.$$

These definitions may be illustrated by figures which show the connection with the idea of area between a curve and its asymptote. Similar definitions would be given if the lower limit were $-\infty$ or if the integrand became infinite at $x=a$. If the integrand were infinite at some intermediate point of the interval, the interval would be subdivided into two intervals and the definition would be applied to each part.

* Here and below the construction of figures is left to the reader.

Now the behavior of $F(x)$ as x approaches a definite value or becomes infinite may be of three distinct sorts; for $F(x)$ may approach a definite finite quantity, or it may become infinite, or it may oscillate without approaching any finite quantity or becoming definitely infinite. The examples

$$\int_0^\infty \frac{8a^3dx}{x^2+4a^2} = \lim_{x=\infty}\left[\int_0^x \frac{8a^3dx}{x^2+4a^2} = 4a^2\tan^{-1}\frac{x}{2a}\right] = 2\pi a^2, \quad \text{a limit,}$$

$$\int_1^\infty \frac{dx}{x} = \lim_{x=\infty}\left[\int_1^x \frac{dx}{x} = \log x\right], \quad \text{becomes infinite, no limit,}$$

$$\int_0^\infty \cos x dx = \lim_{x=\infty}\left[\int_0^x \cos x dx = \sin x\right], \quad \text{oscillates, no limit,}$$

illustrate the three modes of behavior in the case of an infinite upper limit. In the first case, *where the limit exists, the infinite integral is said to converge;* in the other two cases, where the limit does not exist, the integral is said to *diverge.*

If the indefinite integral can be found as above, the question of the convergence or divergence of an infinite integral may be determined and the value of the integral may be obtained in the case of convergence. If the indefinite integral cannot be found, it is of prime importance to know whether the definite infinite integral converges or diverges; for there is little use trying to compute the value of the integral if it does not converge. As the infinite limits or the points where the integrand becomes infinite are the essentials in the discussion of infinite integrals, the integrals will be written with only one limit, as

$$\int^\infty f(x)\,dx, \qquad \int^b f(x)\,dx, \qquad \int_a f(x)\,dx.$$

To discuss a more complicated combination, one would write

$$\int_0^\infty \frac{e^{-x}dx}{\sqrt{x^3}\log x} = \int_0^\xi + \int_\xi^1 + \int_1^\Xi + \int_\Xi^\infty \frac{e^{-x}dx}{\sqrt{x}\log x},$$

and treat all four of the infinite integrals

$$\int_0 \frac{e^{-x}dx}{\sqrt{x^3}\log x}, \quad \int^1 \frac{e^{-x}dx}{\sqrt{x^3}\log x}, \quad \int_1 \frac{e^{-x}dx}{\sqrt{x^3}\log x}, \quad \int^\infty \frac{e^{-x}dx}{\sqrt{x^3}\log x}.$$

Now by definition a function $E(x)$ is called an E-function in the neighborhood of the value $x = a$ when the function is continuous in the neighborhood of $x = a$ and approaches a limit which is neither zero nor infinite (p. 62). *The behavior of the infinite integrals of a function*

which does not change sign and of the product of that function by an E-function are identical as far as convergence or divergence are concerned. Consider the proof of this theorem in a special case, namely,

$$\int^{\infty} f(x)\,dx, \qquad \int^{\infty} f(x)\,E(x)\,dx, \tag{1}$$

where $f(x)$ may be assumed to remain positive for large values of x and $E(x)$ approaches a positive limit as x becomes infinite. Then if K be taken sufficiently large, both $f(x)$ and $E(x)$ have become and will remain positive and finite. By the Theorem of the Mean (Ex. 5, p. 29)

$$m\int_K^x f(x)\,dx < \int_K^x f(x)\,E(x)\,dx < M\int_K^x f(x)\,dx, \qquad x > K,$$

where m and M are the minimum and maximum values of $E(x)$ between K and ∞. Now let x become infinite. As the integrands are positive, the integrals must increase with x. Hence (p. 35)

if $\displaystyle\int_K^{\infty} f(x)\,dx$ converges, $\displaystyle\int_K^{\infty} f(x)\,E(x)\,dx < M\int_K^{\infty} f(x)\,dx$ converges,

if $$\int_K^{\infty} f(x)\,E(x)\,dx \text{ converges,}$$

$$\int_K^{\infty} f(x)\,dx < \frac{1}{m}\int_K^{\infty} f(x)\,E(x)\,dx \text{ converges;}$$

and divergence may be treated in the same way. Hence the integrals (1) converge or diverge together. The same treatment could be given for the case the integrand became infinite and for all the variety of hypotheses which could arise under the theorem.

This theorem is one of the most useful and most easily applied for determining the convergence or divergence of an infinite integral with an integrand which does not change sign. Thus consider the case

$$\int^{\infty} \frac{x\,dx}{(ax+x^2)^{\frac{3}{2}}} = \int^{\infty}\left[\frac{x^2}{ax+x^2}\right]^{\frac{3}{2}}\frac{dx}{x^2}, \quad E(x) = \left[\frac{x^2}{ax+x^2}\right]^{\frac{3}{2}}, \quad \int^{\infty}\frac{dx}{x^2} = -\frac{1}{x}\Big]^{\infty}.$$

Here a simple rearrangement of the integrand throws it into the product of a function $E(x)$, which approaches the limit 1 as x becomes infinite, and a function $1/x^2$, the integration of which is possible. Hence by the theorem the original integral converges. This could have been seen by integrating the original integral; but the integration is not altogether short. Another case, in which the integration is not possible, is

$$\int^1 \frac{dx}{\sqrt{1-x^4}} = \int^1 \frac{1}{\sqrt{1+x^2}\sqrt{1+x}}\frac{dx}{\sqrt{1-x}},$$

$$E(x) = \frac{1}{\sqrt{1+x^2}\sqrt{1+x}}, \qquad \int^1\frac{dx}{\sqrt{1-x}} = -2\sqrt{1-x}\Big]^1.$$

Here $E(1) = \frac{1}{2}$. The integral is again convergent. A case of divergence would be

$$\int_0 \frac{dx}{(2x - x^2)^{\frac{3}{2}}} = \int_0 \frac{1}{(2-x)^{\frac{3}{2}}} \frac{dx}{x^{\frac{3}{2}}}, \qquad E(x) = \frac{1}{(2-x)^{\frac{3}{2}}}, \qquad \int_0 \frac{dx}{x^{\frac{3}{2}}} = -\frac{2}{\sqrt{x}}\bigg|_0.$$

141. The interpretation of a definite integral as an area will suggest another form of test for convergence or divergence in case the integrand does not change sign. Consider two functions $f(x)$ and $\psi(x)$ both of which are, say, positive for large values of x or in the neighborhood of a value of x for which they become infinite. *If the curve $y = \psi(x)$ remains above $y = f(x)$, the integral of $f(x)$ must converge if the integral of $\psi(x)$ converges, and the integral of $\psi(x)$ must diverge if the integral of $f(x)$ diverges.* This may be proved from the definition. For $f(x) < \psi(x)$ and

$$\int_K^x f(x)dx < \int_K^x \psi(x)dx \quad \text{or} \quad F(x) < \Psi(x).$$

Now as x approaches b or ∞, the functions $F(x)$ and $\Psi(x)$ both increase. If $\Psi(x)$ approaches a limit, so must $F(x)$; and if $F(x)$ increases without limit, so must $\Psi(x)$.

As the relative behavior of $f(x)$ and $\psi(x)$ is consequential *only near* particular values of x or when x is very great, the conditions may be expressed in terms of limits, namely: *If $\psi(x)$ does not change sign and if the ratio $f(x)/\psi(x)$ approaches a finite limit (or zero), the integral of $f(x)$ will converge if the integral of $\psi(x)$ converges; and if the ratio $f(x)/\psi(x)$ approaches a finite limit (not zero) or becomes infinite, the integral of $f(x)$ will diverge if the integral of $\psi(x)$ diverges.* For in the first case it is possible to take x so near its limit or so large, as the case may be, that the ratio $f(x)/\psi(x)$ shall be less than any assigned number G greater than its limit; then the functions $f(x)$ and $G\psi(x)$ satisfy the conditions established above, namely $f < G\psi$, and the integral of $f(x)$ converges if that of $\psi(x)$ does. In like manner in the second case it is possible to proceed so far that the ratio $f(x)/\psi(x)$ shall have become to remain greater than any assigned number g less than its limit; then $f > g\psi$, and the result above may be applied to show that the integral of $f(x)$ diverges if that of $\psi(x)$ does.

For an infinite upper limit a direct integration shows that

$$\int^{\infty} \frac{dx}{x^k} = \frac{-1}{k-1}\frac{1}{x^{k-1}}\bigg|^{\infty} \quad \text{or} \quad \log x\bigg|^{\infty}, \qquad \begin{array}{l}\text{converges if } k > 1,\\ \text{diverges if } k \leqq 1.\end{array} \tag{2}$$

Now if the *test function* $\phi(x)$ be chosen as $1/x^k = x^{-k}$, the ratio $f(x)/\phi(x)$ becomes $x^k f(x)$, and *if the limit of the product $x^k f(x)$ exists*

and may be shown to be finite (or zero) as x becomes infinite for any choice of k greater than 1, *the integral of* $f(x)$ *to infinity will converge; but if the product approaches a finite limit (not zero) or becomes infinite for any choice of k less than or equal to* 1, *the integral diverges.* This may be stated as: The integral of $f(x)$ to infinity will converge if $f(x)$ is an infinitesimal of order higher than the first relative to $1/x$ as x becomes infinite, but will diverge if $f(x)$ is an infinitesimal of the first or lower order. In like manner

$$\int^b \frac{dx}{(b-x)^k} = \frac{1}{k-1}\frac{1}{(b-x)^{k-1}}\Bigg|^b \quad \text{or} \quad -\log(b-x)\Bigg|^b, \quad \begin{matrix}\text{converges if } k<1,\\ \text{diverges if } k \geqq 1,\end{matrix} \tag{3}$$

and it may be stated that: The integral of $f(x)$ to b will converge if $f(x)$ is an infinite of order less than the first relative to $(b-x)^{-1}$ as x approaches b, but will diverge if $f(x)$ is an infinite of the first or higher order. The proof is left as an exercise. See also Ex. 3 below.

As an example, let the integral $\int_0^\infty x^n e^{-x}dx$ be tested for convergence or divergence. If $n>0$, the integrand never becomes infinite, and the only integral to examine is that to infinity; but if $n<0$ the integral from 0 has also to be considered. Now the function e^{-x} for large values of x is an infinitesimal of infinite order, that is, the limit of $x^{k+n}e^{-x}$ is zero for any value of k and n. Hence the integrand $x^n e^{-x}$ is an infinitesimal of order higher than the first and the integral to infinity converges under all circumstances. For $x=0$, the function e^{-x} is finite and equal to 1; the order of the infinite $x^n e^{-x}$ will therefore be precisely the order n. Hence the integral from 0 converges when $n>-1$ and diverges when $n \leqq -1$. Hence the function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx, \qquad \alpha > 0,$$

defined by the integral containing the parameter α, will be defined for all positive values of the parameter, but not for negative values nor for 0.

Thus far tests have been established only for integrals in which the integrand does not change sign. There is a general test, not particularly useful for practical purposes, but highly useful in obtaining theoretical results. It will be treated merely for the case of an infinite limit. Let

$$F(x) = \int_K^x f(x)\,dx, \qquad F(x'') - F(x') = \int_{x'}^{x''} f(x)\,dx, \qquad x',\ x'' > K. \tag{4}$$

Now (Ex. 3, p. 44) the necessary and sufficient condition that $F(x)$ approach a limit as x becomes infinite is that $F(x'')-F(x')$ shall approach the limit 0 when x' and x'', regarded as independent variables, become infinite; by the definition, then, this is the necessary and sufficient condition that the integral of $f(x)$ to infinity shall converge. Furthermore

$$\text{if} \qquad \int^{\infty} |f(x)|\, dx \quad \text{converges, then} \int^{\infty} f(x)\, dx \tag{5}$$

must converge and is said to be *absolutely convergent.* The proof of this important theorem is contained in the above and in

$$\int_{x'}^{x''} f(x)\, dx \leqq \int_{x'}^{x''} |f(x)|\, dx.$$

To see whether an integral is absolutely convergent, the tests established for the convergence of an integral with a positive integrand may be applied to the integral of the absolute value, or some obvious direct *method of comparison* may be employed; for example,

$$\int^{\infty} \frac{\cos x dx}{a^2 + x^2} \leqq \int^{\infty} \frac{1\, dx}{a^2 + x^2} \text{ which converges,}$$

and it therefore appears that the integral on the left converges absolutely. When the convergence is not absolute, the question of convergence may sometimes be settled by *integration by parts.* For suppose that the integral may be written as

$$\int^{x} f(x)\, dx = \int^{x} \phi(x)\psi(x)\, dx = \left[\phi(x)\int \psi(x)\, dx\right]^{x} - \int^{x} \phi'(x)\int \psi(x)\, dx^2$$

by separating the integrand into two factors and integrating by parts. Now if, when x becomes infinite, each of the right-hand terms approaches a limit, then

$$\int^{\infty} f(x)\, dx = \lim_{x \doteq \infty}\left[\phi(x)\int \psi(x)\, dx\right]^{x} - \lim_{x \doteq \infty}\int^{x} \phi'(x)\int \psi(x)\, dx dx,$$

and the integral of $f(x)$ to infinity converges.

As an example consider the convergence of $\int^{\infty} \frac{x \cos x dx}{a^2 + x^2}$. Here $\int^{\infty} \frac{x |\cos x| dx}{a^2 + x^2}$ does not appear to be convergent: for, apart from the factor $|\cos x|$ which oscillates between 0 and 1, the integrand is an infinitesimal of only the first order and the integral of such an integrand does not converge; the original integral is therefore apparently not absolutely convergent. However, an integration by parts gives

$$\int^{x} \frac{x \cos x dx}{a^2 + x^2} = \left[\frac{x \sin x}{a^2 + x^2}\right]^{x} - \int^{x} \frac{x^2 - a^2}{(x^2 + a^2)^2} \cos x dx,$$

$$\lim_{x \doteq \infty} \frac{x \sin x}{a^2 + x^2} = 0, \qquad \int^{x} \frac{x^2 - a^2}{(x^2 + a^2)^2} \cos x dx < \int^{x} \frac{dx}{x^2}.$$

Now the integral on the right is seen to be convergent and, in fact, absolutely convergent as x becomes infinite. The original integral therefore must approach a limit and be convergent as x becomes infinite.

EXERCISES

1. Establish the convergence or divergence of these infinite integrals:

(α) $\int^{\infty}\frac{dx}{x\sqrt{1+x^2}}$, (β) $\int^{\infty}\frac{x^2dx}{(a^2+x^2)^2}$, (γ) $\int^{\infty}\frac{x^2dx}{(a^2+x^2)^{\frac{3}{2}}}$,

(δ) $\int_0 x^{\alpha-1}(1-x)^{\beta-1}dx$ (to have an infinite integral, α must be less than 1),

(ϵ) $\int^1 x^{\alpha-1}(1-x)^{\beta-1}dx$, (ζ) $\int_0^a\frac{dx}{\sqrt{ax-x^2}}$, (η) $\int_1^{\infty}\frac{dx}{x\sqrt{x^2-1}}$,

(θ) $\int_0^{\infty}\frac{dx}{1-x^4}$, (ι) $\int_0^2\frac{xdx}{(1-x)^{\frac{1}{3}}}$, (κ) $\int_0^2\frac{x^{\alpha-1}}{1-x}dx$,

(λ) $\int_0^1\frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$, $k<1$, $k=1$, (μ) $\int_0^1\sqrt{\frac{1-k^2x^2}{1-x^2}}dx$, $k<1$.

2. Point out the peculiarities which make these integrals infinite integrals, and test the integrals for convergence or divergence:

(α) $\int_0^1\left(\log\frac{1}{x}\right)^n dx$, conv. if $n>-1$, div. if $n\leqq-1$, (β) $\int_0^1\frac{\log x}{1-x}dx$,

(γ) $\int_0^1(-\log x)^n dx$, (δ) $\int_0^{\frac{\pi}{2}}\log\sin x dx$, (ϵ) $\int_0^{\pi}x\log\sin x dx$,

(ζ) $\int^{\infty}\log\left(x+\frac{1}{x}\right)\frac{dx}{1+x^2}$, (η) $\int_0^{\pi}\frac{dx}{(\sin x+\cos x)^k}$, (θ) $\int_0^1 x^m\left(\log\frac{1}{x}\right)^n dx$,

(ι) $\int_0^{\infty}\frac{e^{-x}dx}{\sqrt{x}\log(x+1)}$, (κ) $\int_0^{\infty}x^{\frac{1}{x}}dx$, (λ) $\int_0^1\log x\tan\frac{\pi x}{2}dx$,

(μ) $\int_0^{\infty}\frac{x^{\alpha-1}}{1+x}dx$, (ν) $\int_{-\infty}^{+\infty}e^{-x^2}dx$, (o) $\int_0^{\infty}\frac{x^{\alpha-1}dx}{(1+x)^2}$,

(π) $\int_0^{\infty}\frac{\sin^2 x}{x^2}dx$, (ρ) $\int_0^1\frac{\log x dx}{\sqrt{1-x^2}}$, (σ) $\int_0^{\infty}e^{-\left(x-\frac{a}{x}\right)^2}$,

(τ) $\int_0^{\infty}\frac{x^{\alpha-1}\log x}{1+x}dx$, (υ) $\int_0^{\infty}\frac{\log(1+a^2x^2)}{1+b^2x^2}dx$, (χ) $\int_0^{\infty}e^{-a^2x^2}\cosh\beta x dx$.

3. Point out the similarities and differences of the method of E-functions and of test functions. Compare also with the work of this section the remark that the determination of the order of an infinitesimal or infinite is a problem in indeterminate forms (p. 63). State also whether it is necessary that $f(x)/\psi(x)$ or $x^kf(x)$ should approach a limit, or whether it is sufficient that the quantity remain finite. Distinguish "of order higher" (p. 356) from "of higher order" (p. 63); see Ex. 8, p. 66.

4. Discuss the convergence of these integrals and prove the convergence is absolute in all cases where possible:

(α) $\int^{\infty}\frac{\sin x}{x^k}dx$, (β) $\int^{\infty}\cos x^2 dx$, (γ) $\int^{\infty}\frac{\cos\sqrt{x}}{x^k}dx$,

(δ) $\int_0^{\infty}\frac{e^{-ax}\sin\beta x}{x}dx$, (ϵ) $\int^{\infty}e^{-a^2x^2}\cos\beta x dx$, (ζ) $\int_0^{\infty}\sqrt{\frac{a^2+x^2}{x^3}}dx$.

(η) $\int_0^\infty \frac{x\sin x}{x^2+k^2}dx$, ($\theta$) $\int_0^\infty e^{-ax}\cos bx dx$, ($\iota$) $\int_0^\infty \frac{\cos x}{\sqrt{x}}dx$,

(κ) $\int_0^\infty x^{\alpha-1}e^{-x\cos\beta}\cos(x\sin\beta)\,dx$, ($\lambda$) $\int_0^\infty \frac{\sin x\cos\alpha x}{x}dx$.

(μ) $\int_0^\infty \cos x^2\cos 2\,\alpha x dx$, ($\nu$) $\int_0^\infty \sin\left(\frac{x^2}{2}+\frac{\alpha^2}{2\,x^2}\right)dx$, ($o$) $\int_0^\infty \frac{\sin^k x^l}{x^m}dx$.

5. If $f_1(x)$ and $f_2(x)$ are two limited functions integrable (in the sense of §§ 28–30) over the integral $a \leqq x \leqq b$, show that their product $f(x)=f_1(x)f_2(x)$ is integrable over the interval. Note that in any interval δ_i, the relations $m_{1i}m_{2i} \leqq m_i \leqq M_i \leqq M_{1i}M_{2i}$ and $M_{1i}M_{2i} - m_{1i}m_{2i} = M_{1i}M_{2i} - M_{1i}m_{2i} + M_{1i}m_{2i} - m_{1i}m_{2i} = M_{1i}O_{2i} + m_{2i}O_{1i}$ hold. Show further that

$$\int_a^b f_1(x)f_2(x)\,dx = \lim\sum f_1(\xi_i)f_2(\xi_i)\delta_i$$
$$= \lim\sum f_1(\xi_i)\left[\int_{x_i}^{x_{i+1}} f_2(x)\,dx - \int_{x_i}^{x_{i+1}}\{f_2(\xi_i) - f_2(x)\,dx\}\right],$$

or
$$\int_a^b f(x)\,dx = \lim\sum f_1(\xi_i)\int_{x_i}^{x_{i+1}} f_2(x)\,dx$$
$$= \lim\sum f_1(\xi_i)\left[\int_{x_i}^{b} f_2(x)\,dx - \int_{x_{i+1}}^{b} f_2(x)\,dx\right],$$

or
$$\int_a^b f(x)\,dx = f_1(\xi_1)\int_a^b f_2(x)\,dx + \lim\sum\,[f_2(\xi_i) - f_2(\xi_{i-1})]\int_{x_i}^b f_2(x)\,dx.$$

6. *The Second Theorem of the Mean.* If $f(x)$ and $\phi(x)$ are two limited functions integrable in the interval $a \leqq x \leqq b$, and if $\phi(x)$ is positive, nondecreasing, and less than K, then

$$\int_a^b \phi(x)f(x)\,dx - K\int_\xi^b f(x)\,dx, \qquad a \leqq \xi \leqq b.$$

And, more generally, if $\phi(x)$ satisfies $-\infty < k \leqq \phi(x) \leqq K < \infty$ and is either nondecreasing or nonincreasing throughout the interval, then

$$\int_a^b \phi(x)f(x)\,dx = k\int_a^\xi f(x)\,dx + K\int_\xi^b f(x)\,dx. \qquad a \leqq \xi \leqq b.$$

In the first case the proof follows from Ex. 5 by noting that the integral of $\phi(x)f(x)$ may be regarded as the limit of the sum

$$\phi(\xi_1)\int_a^b f(x)\,dx + \sum\,[\phi(\xi_i) - \phi(\xi_{i-1})]\int_{x_i}^b f(x)\,dx + [K - \phi(\xi_n)]\int_{x_n}^b f(x)\,dx,$$

where the restrictions on $\phi(x)$ make the coefficients of the integrals all positive or zero, and where the sum may consequently be written as

$$\mu\,[\phi(\xi_1) + \phi(\xi_2) - \phi(\xi_1) + \cdots + \phi(\xi_n) - \phi(\xi_{n-1}) + K - \phi(\xi_n)] = \mu K$$

if μ be a properly chosen mean value of the integrals which multiply these coefficients; as the integrals are of the form $\int_\xi^b f(x)\,dx$ where $\xi = a, x_1, \cdots, x_n$, it follows

that μ must be of the same form where $a \leqq \xi \leqq b$. The second form of the theorem follows by considering the function $\phi - k$ or $k - \phi$.

7. If $\phi(x)$ is a function varying always in the same sense and approaching a finite limit as x becomes infinite, the integral $\int^{\infty} \phi(x) f(x)\, dx$ will converge if $\int^{\infty} f(x)\, dx$ converges. Consider

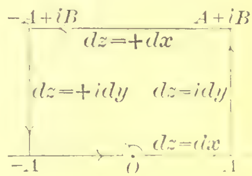
$$\int_{x'}^{x''} \phi(x) f(x)\, dx = \phi(x') \int_{x'}^{\xi} f(x)\, dx + \phi(x'') \int_{\xi}^{x''} f(x)\, dx.$$

8. If $\phi(x)$ is a function varying always in the same sense and approaching 0 as a limit when $x = \infty$, and if the integral $F(x)$ of $f(x)$ remains finite when $x = \infty$, then the integral $\int^{\infty} \phi(x) f(x)\, dx$ is convergent. Consider

$$\int_{x'}^{x''} \phi(x) f(x)\, dx = \phi(x') [F(\xi) - F(x')] + \phi(x'') [F(x'') - F(\xi)].$$

This test is very useful in practice; for many integrals are of the form $\int^{\infty} \phi(x) \sin x dx$ where $\phi(x)$ constantly decreases or increases toward the limit 0 when $x = \infty$; all these integrals converge.

142. The evaluation of infinite integrals. After an infinite integral has been proved to converge, the problem of calculating its value still remains. No general method is to be had, and for each integral some special device has to be discovered which will lead to the desired result. *This may frequently be accomplished by choosing a function $F(z)$ of the complex variable $z = x + iy$ and integrating the function around some closed path in the z-plane.* It is known that if the points where $F(z) = X(x, y) + iY(x, y)$ ceases to have a derivative $F'(z)$, that is, where $X(x, y)$ and $Y(x, y)$ cease to have continuous first partial derivatives satisfying the relations $X'_x = Y'_y$ and $X'_y = -Y'_x$, are cut out of the plane, the integral of $F(z)$ around any closed path which does not include any of the excised points is zero (§ 124). It is sometimes possible to select such a function $F(z)$ and such a path of integration that part of the integral of the complex function reduces to the given infinite integral while the rest of the integral of the complex function may be computed. Thus there arises an equation which determines the value of the infinite integral.



Consider the integral $\int_0^{\infty} \frac{\sin x}{x} dx$ which is known to converge. Now

$$\int_0^{\infty} \frac{\sin x}{x} dx = \int_0^{\infty} \frac{e^{ix} - e^{-ix}}{2ix} dx = \int_0^{\infty} \frac{e^{ix}}{2ix} - \int_0^{\infty} \frac{e^{-ix}}{2ix} dx$$

suggests at once that the function e^{iz}/z be examined. This function has a definite derivative at every point except $z = 0$, and the origin is therefore the only point

which has to be cut out of the plane. The integral of e^{iz}/z around any path such as that marked in the figure * is therefore zero. Then if a is small and A is large,

$$0=\int_{\smile}\frac{e^{iz}}{z}dz=\int_a^A\frac{e^{ix}}{x}dx+\int_0^B\frac{e^{iA-y}}{A+iy}i\,dy+\int_A^{-A}\frac{e^{ix-B}}{x+iB}dx$$
$$+\int_B^0\frac{e^{-iA-y}}{-A+iy}i\,dy+\int_{-A}^{-a}\frac{e^{ix}}{x}dx+\int_{-a}^{+a}\frac{e^{iz}}{z}dz.$$

But $\displaystyle\int_{-A}^{-a}\frac{e^{ix}}{x}dx=-\int_{-a}^{-A}\frac{e^{ix}}{x}dx=-\int_a^A\frac{e^{-ix}dx}{x}$ and $\displaystyle\int_{-a}^{+a}\frac{e^{iz}}{z}dz=\int_{-a}^{+a}\frac{1+\eta}{z}dz$:

the first by the ordinary rules of integration and the second by Maclaurin's Formula. Hence

$$0=\int_{\smile}\frac{e^{iz}}{z}dz=\int_a^A\frac{e^{ix}-e^{-ix}}{x}+\int_{-a}^{+a}\frac{dz}{z}+\text{four other integrals.}$$

It will now be shown that by taking the rectangle sufficiently large and the semicircle about the origin sufficiently small each of the four integrals may be made as small as desired. The method is to replace each integral by a larger one which may be evaluated.

$$\left|\int_0^B\frac{e^{iA-y}}{A+iy}i\,dy\right|\leqq\int_0^B\frac{|e^{iA}|\,|e^{-y}|}{|A+iy|}|i|\,|dy|<\int_0^B\frac{1}{A}e^{-y}dy<\frac{B}{A}.$$

These changes involve the facts that the integral of the absolute value is as great as the absolute value of the integral and that $e^{iA-y}=e^{iA}e^{-y}$, $|e^{iA}|=1$, $|A+iy|>A$, $e^{-y}<1$. For the relations $|e^{iA}|=1$ and $|A+iy|>A$, the interpretation of the quantities as vectors suffices (§§ 71–74); that the integral of the absolute value is as great as the absolute value of the integral follows from the same fact for a sum (p. 154). The absolute value of a fraction is enlarged if that of its numerator is enlarged or that of its denominator diminished. In a similar manner

$$\left|\int_A^{-A}\frac{e^{ix-B}}{x+iB}dx\right|<\int_{-A}^{A}\frac{e^{-B}}{B}dx=2e^{-B}\frac{A}{B},\qquad\left|\int_B^0\frac{e^{-iA-y}}{-A+iy}i\,dy\right|<\frac{B}{A}.$$

Furthermore $$\left|\int_{-a}^{+a}\frac{\eta}{z}dz\right|\leqq\int_{-a}^{+a}|\eta|\frac{|dz|}{|z|}=\int_0^\pi|\eta|\,d\phi.$$

$$\int_{-a}^{+a}\frac{dz}{z}=\int_\pi^0\frac{re^{\phi i}i\,d\phi}{re^{\phi i}}=-\pi i.$$

Then $\displaystyle 0=\int_{\smile}\frac{e^{iz}}{z}dz=\int_a^A 2i\frac{\sin x}{x}dx-\pi i+R,\qquad |R|<2\frac{B}{A}+2e^{-B}\frac{A}{B}+\pi\epsilon,$

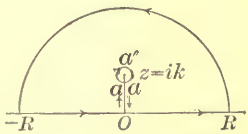
where ϵ is the greatest value of $|\eta|$ on the semicircle. Now let the rectangle be so chosen that $A=Be^{\frac{1}{2}B}$; then $|R|<4e^{-\frac{1}{2}B}+\pi\epsilon$. By taking B sufficiently large $e^{-\frac{1}{2}B}$ may be made as small as desired; and by taking the semicircle sufficiently

* It is also possible to integrate along a semicircle from A to $-A$, or to come back directly from iB to the origin and separate real from imaginary parts. These variations in method may be left as exercises.

small, ϵ may be made as small as desired. This amounts to saying that, for A sufficiently large and for a sufficiently small, R is negligible. In other words, by taking A large enough and a small enough $\int_a^A \frac{\sin x}{x}$ may be made to differ from $\frac{\pi}{2}$ by as little as desired. As the integral from zero to infinity converges and may be regarded as the limit of the integral from a to A (is so defined, in fact), the integral from zero to infinity must also differ from $\frac{1}{2}\pi$ by as little as desired. But if two constants differ from each other by as little as desired, they must be equal. Hence

$$\int_0^\infty \frac{\sin x}{x} = \frac{\pi}{2}. \tag{6}$$

As a second example consider what may be had by integrating $e^{iz}/(z^2+k^2)$ over an appropriate path. The denominator will vanish when $z = \pm ik$ and there are two points to exclude in the z-plane. Let the integral be extended over the closed path as indicated. There is no need of integrating back and forth along the double line Oa, because the function takes on the same values and the integrals destroy each other. Along the large semicircle $z = Re^{i\phi}$ and $dz = Rie^{i\phi}d\phi$. Moreover



$$\int_{-R}^0 \frac{e^{ix}dx}{x^2+k^2} = -\int_0^{-R} \frac{e^{ix}dx}{x^2+k^2} = \int_0^R \frac{e^{-ix}dx}{x^2+k^2} \qquad \text{by elementary rules.}$$

Hence $$\int_{-R}^0 \frac{e^{ix}dx}{x^2+k^2} + \int_0^R \frac{e^{ix}dx}{x^2+k^2} = \int_0^R \frac{e^{ix}+e^{-ix}}{x^2+k^2}dx = 2\int_0^R \frac{\cos x}{x^2+k^2}dx,$$

and $$0 = \int_\bigcirc \frac{e^{iz}}{z^2+k^2}dz = 2\int_0^R \frac{\cos x}{x^2+k^2}dx + \int_0^\pi \frac{e^{iRe^{i\phi}}Rie^{i\phi}d\phi}{R^2e^{2i\phi}+k^2} + \int_{aa'a} \frac{e^{iz}dz}{z^2+k^2}.$$

Now $$|e^{iRe^{i\phi}}| = |e^{iR(\cos\phi + i\sin\phi)}| = |e^{-R\sin\phi}e^{iR\cos\phi}| = e^{-R\sin\phi}.$$

Moreover $|R^2e^{2i\phi}+k^2|$ cannot possibly exceed R^2-k^2 and can equal it only when $\phi = \frac{1}{2}\pi$. Hence

$$\left|\int_0^\pi \frac{e^{iRe^{i\phi}}Rie^{i\phi}d\phi}{R^2e^{2i\phi}+k^2}\right| \leqq \int_0^\pi \frac{Re^{-R\sin\phi}}{R^2-k^2}d\phi = 2\int_0^{\frac{\pi}{2}} \frac{Re^{-R\sin\phi}}{R^2-k^2}d\phi.$$

Now by Ex. 28, p. 11, $\sin\phi > 2\phi/\pi$. Hence the integral may be further increased.

$$\left|\int_0^\pi \frac{e^{iRe^{i\phi}}Rie^{i\phi}d\phi}{R^2e^{2i\phi}+k^2}\right| < 2\int_0^{\frac{\pi}{2}} \frac{Re^{-R\frac{2\phi}{\pi}}d\phi}{R^2-k^2} = \frac{\pi}{R^2-k^2}(e^{-R}-1).$$

Moreover, $$\int_{aa'a} \frac{e^{iz}dz}{z^2+k^2} = \int_{aa'a} \frac{e^{iz}}{z+ik}\frac{dz}{z-ik} = \int_{aa'a}\left(\frac{e^{-k}}{2ki}+\eta\right)\frac{dz}{z-ik},$$

where η is uniformly infinitesimal with the radius of the small circle. But

$$\int_{aa'a} \frac{dz}{z-ik} = 2\pi i, \quad \text{and} \quad \int_{aa'a} \frac{e^{iz}dz}{z^2+k^2} = \frac{2\pi e^{-k}}{2k} + \zeta,$$

where $|\zeta| \leqq 2\pi\epsilon$ if ϵ is the largest value of $|\eta|$. Hence finally

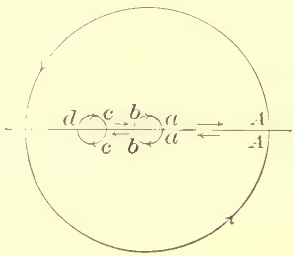
$$0 = 2\int_0^R \frac{\cos x}{x^2 + k^2}\,dx - \frac{\pi}{k}e^{-k} + \zeta + \frac{\pi}{R^2 - k^2}(e^{-R} - 1).$$

By taking the small circle small enough and the large circle large enough, the last two terms may be made as near zero as desired. Hence

$$\int_0^\infty \frac{\cos x}{x^2 + k^2}\,dx = \frac{\pi e^{-k}}{2k}. \tag{7}$$

It may be noted that, by the work of § 126, $\int_{aa'a} \frac{e^{iz}}{z + ki}\frac{dz}{z - ki} = -2\pi i\frac{e^{-k}}{2ki}$ is exact and not merely approximate, and remains exact for any closed curve about $z = ki$ which does not include $z = -ki$. That it is approximate in the small circle follows immediately from the continuity of $e^{iz}/(z + ki) = e^{-k}/2ki + \eta$ and a direct integration about the circle.

As a third example of the method let $\int_0^\infty \frac{x^{\alpha - 1}}{1 + x}dx$ be evaluated. This integral will converge if $0 < \alpha < 1$, because the infinity at the origin is then of order less than the first and the integrand is an infinitesimal of order higher than the first for large values of x. The function $z^{\alpha-1}/(1 + z)$ becomes infinite at $z = 0$ and $z = -1$, and these points must be excluded. The path marked in the figure is a closed path which does not contain them. Now here the integral back and forth along the line aA cannot be neglected; for the function has a fractional or irrational power $z^{\alpha-1}$ in the numerator and is therefore not single valued. In fact, when z is given, the function $z^{\alpha-1}$ is determined as far as its absolute value is concerned, but its angle may take on any addition of the form $2\pi k(\alpha - 1)$ with k integral. Whatever value of the function is assumed at one point of the path, the values at the other points must be such as to piece on continuously when the path is followed. Thus the values along the line aA outward will differ by $2\pi(\alpha - 1)$ from those along Aa inward because the turn has been made about the origin and the angle of z has increased by 2π. The double line bc and cb, however, may be disregarded because no turn about the origin is made in describing cdc. Hence, remembering that $e^{\pi i} = -1$,



$$0 = \int_{\bigcirc}\frac{z^{\alpha-1}}{1+z}dz = \int_{\bigcirc}\frac{r^{\alpha-1}e^{(\alpha-1)\phi i}}{1 + re^{\phi i}}d(re^{\phi i}) = \int_a^A \frac{r^{\alpha-1}}{1+r}dr + \int_0^{2\pi}\frac{A^\alpha e^{\alpha\phi i}}{1 + Ae^{\phi i}}i\,d\phi$$

$$+ \int_A^a \frac{r^{\alpha-1}e^{2\pi(\alpha-1)i}}{1 + re^{2\pi i}}e^{2\pi i}dr + \int_{abba}\frac{z^{\alpha-1}}{1+z}dz + \int_{cdc}\frac{z^{\alpha-1}}{1+z}dz.$$

Now
$$\int_a^A \frac{r^{\alpha-1}}{1+r}dr + \int_A^a \frac{r^{\alpha-1}e^{2\pi\alpha i}}{1+r}dr = \int_a^A \frac{r^{\alpha-1}}{1+r}(1 - e^{2\pi\alpha i})\,dr,$$

$$\left|\int_0^{2\pi}\frac{A^\alpha e^{\alpha\phi i}}{1 + Ae^{\phi i}}i\,d\phi\right| \leqq \int_0^{2\pi}\frac{A^\alpha}{A-1}\left|e^{\alpha\phi i}\right|d\phi = \frac{2\pi A^\alpha}{A-1},$$

$$\left|\int_{abba}\frac{z^{\alpha-1}}{1+z}dz\right| = \left|\int_{2\pi}^{0}\frac{a^\alpha e^{\alpha\phi i}}{1 + ae^{\phi i}}i\,d\phi\right| \leqq \int_0^{2\pi}\frac{a^\alpha}{1-a}d\phi = \frac{2\pi a^\alpha}{1-a},$$

$$\int_{\text{circle}} \frac{z^{\alpha-1}}{1+z} dz = \int z^{\alpha-1} \frac{dz}{1+z} = -2\pi i(-1)^{\alpha-1} = -2\pi i e^{\pi(\alpha-1)i} = 2\pi i e^{\pi\alpha i}.$$

Hence $\quad 0 = (1 - e^{2\pi\alpha i}) \displaystyle\int_a^A \frac{r^{\alpha-1}}{1+r} dr + 2\pi i e^{\pi\alpha i} + \zeta, \qquad |\zeta| < \frac{2\pi A^{\alpha}}{A-1} + \frac{2\pi a^{\alpha}}{1-a}.$

If A be taken sufficiently large and a sufficiently small, ζ may be made as small as desired. Then by the same reasoning as before it follows that

$$0 = (1 - e^{2\pi\alpha i}) \int_0^{\infty} \frac{r^{\alpha-1}}{1+r} dr + 2\pi i e^{\pi\alpha i}, \quad \text{or} \quad 0 = -\sin \pi\alpha \int_0^{\infty} \frac{r^{\alpha-1}}{1+r} dr + \pi,$$

and
$$\int_0^{\infty} \frac{x^{\alpha-1}}{1+x} dx = \frac{\pi}{\sin \alpha\pi}. \tag{8}$$

143. One integral of particular importance is $\displaystyle\int_0^{\infty} e^{-x^2} dx$. The evaluation may be made by a device which is rarely useful. Write

$$\int_0^A e^{-x^2} dx = \left[\int_0^A e^{-x^2} dx \int_0^A e^{-y^2} dy\right]^{\frac{1}{2}} = \left[\int_0^A \int_0^A e^{-x^2-y^2} dx dy\right]^{\frac{1}{2}}.$$

The passage from the product of two integrals to the double integral may be made because neither the limits nor the integrands of either integral depend on the variable in the other. Now transform to polar coördinates and integrate over a quadrant of radius A.

$$\int_0^A \int_0^A e^{-x^2-y^2} dx dy = \int_0^{\frac{\pi}{2}} \int_0^A e^{-r^2} r dr d\theta + R = \frac{1}{4}\pi(1 - e^{-A^2}) + R,$$

where R denotes the integral over the area between the quadrant and square, an area less than $\frac{1}{2}A^2$ over which $e^{-r^2} \leqq e^{-A^2}$. Then

$$R < \tfrac{1}{2} A^2 e^{-A^2}, \qquad \left| \int_0^A \int_0^A e^{-x^2-y^2} dx dy - \tfrac{1}{4}\pi \right| < \tfrac{1}{2} A^2 e^{-A^2}.$$

Now A may be taken so large that the double integral differs from $\frac{1}{4}\pi$ by as little as desired, and hence for sufficiently large values of A the simple integral will differ from $\frac{1}{2}\sqrt{\pi}$ by as little as desired. Hence *

$$\int_0^{\infty} e^{-x^2} dx = \tfrac{1}{2}\sqrt{\pi}. \tag{9}$$

* It should be noticed that the proof just given does not require the theory of infinite double integrals nor of change of variable; the whole proof consists merely in finding a number $\frac{1}{2}\sqrt{\pi}$ from which the integral may be shown to differ by as little as desired. This was also true of the proofs in § 142; no theory had to be developed and no limiting processes were used. In fact the evaluations that have been performed show of themselves that the infinite integrals converge. For when it has been shown that an integral with a large enough upper limit and a small enough lower limit can be made to differ from a certain constant by as little as desired, it has thereby been proved that that integral from zero to infinity must converge to the value of that constant.

When some infinite integrals have been evaluated, others may be obtained from them by various operations, such as integration by parts and change of variable. It should, however, be borne in mind that the rules for operating with definite integrals were established only for finite integrals and must be *reëstablished* for infinite integrals. From the direct application of the definition it follows that the integral of a function times a constant is the product of the constant by the integral of the function, and that the sum of the integrals of two functions taken between the same limits is the integral of the sum of the functions. But it cannot be inferred conversely that an integral may be resolved into a sum as

$$\int_a^b [f(x)+\phi(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b \phi(x)\,dx$$

when one of the limits is infinite or one of the functions becomes infinite in the interval. For, the fact that the integral on the left converges is no guarantee that either integral upon the right will converge; all that can be stated is that *if one of the integrals on the right converges, the other will*, and the equation will be true. The same remark applies to integration by parts,

$$\int_a^b f(x)\,\phi'(x)\,dx = \Big[f(x)\,\phi(x)\Big]_a^b - \int_a^b f'(x)\,\phi(x)\,dx.$$

If, in the process of taking the limit which is required in the definition of infinite integrals, *two of the three terms in the equation approach limits, the third will approach a limit*, and the equation will be true for the infinite integrals.

The formula for the change of variable is

$$\int_{x=\phi(t)}^{X=\phi(T)} f(x)\,dx = \int_t^T f[\phi(t)]\,\phi'(t)\,dt,$$

where it is assumed that the derivative $\phi'(t)$ is continuous and does not vanish in the interval from t to T (although either of these conditions may be violated at the extremities of the interval). As these two quantities are equal, they will approach equal limits, provided they approach limits at all, when the limit

$$\int_{a=\phi(t_0)}^{b=\phi(t_1)} f(x)\,dx = \int_{t_0}^{t_1} f[\phi(t)]\,\phi'(t)\,dt$$

required in the definition of an infinite integral is taken, where one of the four limits a, b, t_0, t_1 is infinite or one of the integrands becomes

infinite at the extremity of the interval. *The formula for the change of variable is therefore applicable to infinite integrals.* It should be noted that the proof applies only to infinite limits and infinite values of the integrand at the extremities of the interval of integration; in case the integrand becomes infinite within the interval, the change of variable should be examined in each subinterval just as the question of convergence was examined.

As an example of the change of variable consider $\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}$ and take $x = \alpha x'$.

$$\int_{x=0}^{x=\infty} \frac{\sin \alpha x'}{x'} dx' = \int_{x'=0}^{+\infty} \frac{\sin \alpha x'}{x'} dx' \quad \text{or} \quad = \int_{x'=0}^{-\infty} \frac{\sin \alpha x'}{x'} dx' = -\int_{x'=0}^{x'=+\infty} \frac{\sin \alpha x'}{x'} dx',$$

according as α is positive or negative. Hence the results

$$\int_0^\infty \frac{\sin \alpha x}{x} dx = +\frac{\pi}{2} \quad \text{if} \quad \alpha > 0 \quad \text{and} \quad -\frac{\pi}{2} \quad \text{if} \quad \alpha < 0. \tag{10}$$

Sometimes changes of variable or integrations by parts will lead back to a given integral in such a way that its value may be found. For instance take

$$I = \int_0^{\frac{\pi}{2}} \log \sin x dx = -\int_{\frac{\pi}{2}}^0 \log \cos y dy = \int_0^{\frac{\pi}{2}} \log \cos y dy, \qquad y = \frac{\pi}{2} - x.$$

Then
$$2I = \int_0^{\frac{\pi}{2}} (\log \sin x + \log \cos x) dx = \int_0^{\frac{\pi}{2}} \log \frac{\sin 2x}{2} dx$$

$$= \frac{1}{2}\int_0^{\pi} \log \sin x dx - \frac{\pi}{2} \log 2 = \int_0^{\frac{\pi}{2}} \log \sin x dx - \frac{\pi}{2} \log 2.$$

Hence
$$I = \int_0^{\frac{\pi}{2}} \log \sin x dx = -\frac{\pi}{2} \log 2. \tag{11}$$

Here the first change was $y = \frac{1}{2}\pi - x$. The new integral and the original one were then added together (the variable indicated under the sign of a definite integral is immaterial, p. 26), and the sum led back to the original integral by virtue of the substitution $y = 2x$ and the fact that the curve $y = \log \sin x$ is symmetrical with respect to $x = \frac{1}{2}\pi$. This gave an equation which could be solved for I.

EXERCISES

1. Integrate $\frac{ze^{iz}}{z^2 + k^2}$, as for the case of (7), to show $\int_0^\infty \frac{x \sin x}{x^2 + k^2} dx = \frac{\pi}{2} e^{-k}$.

2. By direct integration show that $\int_0^\infty e^{-(a-bi)z} dz$ converges to $(a - bi)^{-1}$, when $a > 0$ and the integral is extended along the line $y = 0$. Thus prove the relations

$$\int_0^\infty e^{-ax} \cos bx dx = \frac{a}{a^2 + b^2}, \qquad \int_0^\infty e^{-ax} \sin bx dx = \frac{b}{a^2 + b^2}, \qquad a > 0.$$

Along what lines issuing from the origin would the given integral converge?

3. Show $\int_0^\infty \frac{x^{\alpha-1}dx}{(1+x)^2} = \frac{(1-\alpha)\pi}{\sin\alpha\pi}$. To integrate about $z = -1$ use the binomial expansion $z^{\alpha-1} = [-1+1+z]^{\alpha-1} = (-1)^{\alpha-1}[1+(1-\alpha)(1+z)+\eta(1+z)]$, η small.

4. Integrate e^{-z^2} around a circular sector with vertex at $z = 0$ and bounded by the real axis and a line inclined to it at an angle of $\frac{1}{4}\pi$. Hence show

$$e^{\frac{1}{4}\pi i}\int_0^\infty (\cos r^2 - i\sin r^2)\,dr = \int_0^\infty e^{-x^2}dx = \frac{\sqrt{\pi}}{2},$$

$$\int_0^\infty \cos x^2 dx = \int_0^\infty \sin x^2 dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}.$$

5. Integrate e^{-z^2} around a rectangle $y = 0$, $y = B$, $x = \pm A$, and show

$$\int_0^\infty e^{-x^2}\cos 2\,ax dx = \tfrac{1}{2}\sqrt{\pi}e^{-a^2}, \qquad \int_{-\infty}^\infty e^{-x^2}\sin 2\,axdx = 0.$$

6. Integrate $z^{\alpha-1}e^{-z}$, $0<\alpha$, along a sector of angle $q < \frac{1}{2}\pi$ to show

$$\sec\alpha q\int_0^\infty x^{\alpha-1}e^{-x\cos q}\cos(x\sin q)\,dx$$
$$= \csc\alpha q\int_0^\infty x^{\alpha-1}e^{-x\cos q}\sin(x\sin q)\,dx = \int_0^\infty x^{\alpha-1}e^{-x}dx.$$

7. Establish the following results by the proper change of variable:

(α) $\int_0^\infty \frac{\cos\alpha x}{x^2+k^2}dx = \frac{\pi e^{-ak}}{2k}$, $\alpha>0$, $\qquad (\beta)$ $\int_0^\infty \frac{x^{\alpha-1}dx}{\beta+x} = \frac{\pi\beta^{\alpha-1}}{\sin\alpha\pi}$, $\beta>0$,

(γ) $\int_0^\infty e^{-\alpha^2x^2}dx = \frac{1}{2\alpha}\sqrt{\pi}$, $\qquad (\delta)$ $\int_0^\infty e^{-\alpha x}\frac{1}{\sqrt{x}}dx = \sqrt{\frac{\pi}{\alpha}}$,

(ϵ) $\int_0^\infty e^{-\alpha^2x^2}\cos bxdx = \frac{\sqrt{\pi}e^{-\frac{b^2}{4a^2}}}{2\alpha}$, $\alpha>0$, $\qquad (\zeta)$ $\int_0^1 \frac{dx}{\sqrt{-\log x}} = \sqrt{\pi}$,

(η) $\int_0^\infty \frac{\cos x}{\sqrt{x}}dx = \int_0^\infty \frac{\sin x}{\sqrt{x}}dx = \sqrt{\frac{\pi}{2}}$, $\qquad (\theta)$ $\int_0^1 \frac{\log xdx}{\sqrt{1-x^2}} = -\frac{\pi}{2}\log 2$.

8. By integration by parts or other devices show the following:

(α) $\int_0^\pi x\log\sin xdx = -\frac{1}{2}\pi^2\log 2$, $\qquad (\beta)$ $\int_0^\infty \frac{\sin^2 x}{x^2}dx = \frac{\pi}{2}$,

(γ) $\int_0^\infty \frac{\sin x\cos\alpha x}{x}dx = \frac{\pi}{2}$ if $-1<\alpha<1$, or $\frac{\pi}{4}$ if $\alpha = \pm 1$, or 0 if $|\alpha|>1$,

(δ) $\int_0^\infty x^2e^{-\alpha^2x^2}dx = \frac{\sqrt{\pi}}{4\alpha^3}$, $\qquad (\epsilon)$ $\int_0^\infty x^4e^{-\alpha^2x^2}dx = \frac{3\sqrt{\pi}}{8\alpha^5}$,

(ζ) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ if $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$, $\qquad (\eta)$ $\int_0^\pi \frac{x\sin xdx}{1+\cos^2 x} = \frac{\pi^2}{4}$,

(θ) $\int_0^\infty \log\left(x+\frac{1}{x}\right)\frac{dx}{1+x^2} = \pi\log 2$, by virtue of $x = \tan y$.

9. Suppose $\int_a^\infty f(x)\frac{dx}{x}$, where $a>0$, converges. Then if $p>0$, $q>0$,

$$\int_0^\infty \frac{f(px)-f(qx)}{x}dx = \lim_{a\doteq 0}\left[\int_a^\infty \frac{f(px)-f(qx)}{x}dx = \int_{pa}^\infty \frac{f(\xi)}{\xi}d\xi - \int_{qa}^\infty \frac{f(\xi)}{\xi}d\xi\right].$$

Show
$$\int_0^\infty \frac{f(px)-f(qx)}{x}dx = \lim_{a\doteq 0}\int_{pa}^{qa} f(x)\frac{dx}{x} = f(0)\log\frac{q}{p}.$$

Hence (α) $\displaystyle\int_0^\infty \frac{\sin px - \sin qx}{x}dx = 0.$ $\qquad(\beta)$ $\displaystyle\int_0^\infty \frac{e^{-px}-e^{-qx}}{x}dx = \log\frac{q}{p},$

(γ) $\displaystyle\int_0^1 \frac{x^{p-1}-x^{q-1}}{\log x}dx = \log\frac{q}{p},$ $\qquad(\delta)$ $\displaystyle\int_0^\infty \frac{\cos x - \cos ax}{x}dx = \log a.$

10. If $f(x)$ and $f'(x)$ are continuous, show by integration by parts that

$$\lim_{k=\infty}\int_a^b f(x)\sin kx\,dx = 0. \quad \text{Hence prove} \quad \lim_{k=\infty}\int_0^a f(x)\frac{\sin kx}{x}dx = \frac{\pi}{2}f(0).$$

$$\left[\text{Write}\quad \int_0^a f(x)\frac{\sin kx}{x}dx = f(0)\int_0^a \frac{\sin kx}{x}dx + \int_0^a \frac{f(x)-f(0)}{x}\sin kx\,dx.\right]$$

Apply Ex. 6, p. 359, to prove these formulas under general hypotheses.

11. Show that $\displaystyle\lim_{k=\infty}\int_a^b f(x)\frac{\sin kx}{x}dx = 0$ if $b>a>0$. Hence note that

$$\lim_{k=\infty}\lim_{a=0}\int_a^b f(x)\frac{\sin kx}{x}dx \neq \lim_{a=0}\lim_{k=\infty}\int_a^b f(x)\frac{\sin kx}{x}dx, \quad \text{unless} \quad f(0)=0.$$

144. Functions defined by infinite integrals. If the integrand of an integral contains a parameter (§ 118), the integral defines a function of the parameter for every value of the parameter for which it converges. The continuity and the differentiability and integrability of the function have to be treated. Consider first the case of an infinite limit

$$\int_a^\infty f(x,\alpha)dx = \int_a^x f(x,\alpha)dx + R(x,\alpha), \qquad R = \int_x^\infty f(x,\alpha)dx.$$

If this integral is to converge for a given value $\alpha=\alpha_0$, it is necessary that the remainder $R(x,\alpha_0)$ can be made as small as desired by taking x large enough, and shall remain so for all larger values of x. In like manner if the integrand becomes infinite for the value $x=b$, the condition that

$$\int_a^b f(x,\alpha)dx = \int_a^x f(x,\alpha)dx + R(x,\alpha), \qquad R = \int_x^b f(x,\alpha)dx$$

converge is that $R(x,\alpha_0)$ can be made as small as desired by taking x near enough to b, and shall remain so for nearer values.

Now for different values of α, the least values of x which will make $|R(x,\alpha)| \leqq \epsilon$, when ϵ is assigned, will probably differ. The infinite integrals are said to *converge uniformly* for a range of values of α such as

$\alpha_0 \leqq \alpha \leqq \alpha_1$ when it is possible to take x so large (or x so near b) that $|R(x, \alpha)| < \epsilon$ holds (and continues to hold for all larger values, or values nearer b) simultaneously for all values of α in the range $\alpha_0 \leqq \alpha \leqq \alpha_1$. The most useful test for uniform convergence is contained in the theorem: *If a positive function $\phi(x)$ can be found such that*

$$\int^{\infty} \phi(x)\,dx \quad \textit{converges and} \quad \phi(x) \geqq |f(x, \alpha)|$$

for all large values of x and for all values of α in the interval $\alpha_0 \leqq \alpha \leqq \alpha_1$, the integral of $f(x, \alpha)$ to infinity converges uniformly (and absolutely) for the range of values in α. The proof is contained in the relation

$$\left|\int_x^{\infty} f(x, \alpha)\,dx\right| \leqq \int_x^{\infty} \phi(x)\,dx < \epsilon,$$

which holds for all values of α in the range. There is clearly a similar theorem for the case of an infinite integrand. See also Ex. 18 below.

Fundamental theorems are:* Over any interval $\alpha_0 \leqq \alpha \leqq \alpha_1$ where an infinite integral converges uniformly the integral defines a continuous function of α. This function may be integrated over any finite interval where the convergence is uniform by integrating with respect to α under the sign of integration with respect to x. The function may be differentiated at any point α_ξ of the interval $\alpha_0 \leqq \alpha \leqq \alpha_1$ by differentiating with respect to α under the sign of integration with respect to x provided the integral obtained by this differentiation converges uniformly for values of α in the neighborhood of α_ξ. Proofs of these theorems are given immediately below.†

To prove that the function is continuous if the convergence is uniform let

$$\psi(\alpha) = \int_a^{\infty} f(x, \alpha)\,dx = \int_a^x f(x, \alpha)\,dx + R(x, \alpha), \qquad \alpha_0 \leqq \alpha \leqq \alpha_1,$$

$$\psi(\alpha + \Delta\alpha) = \int_a^x f(x, \alpha + \Delta\alpha)\,dx + R(x, \alpha + \Delta\alpha),$$

$$|\Delta\psi| \leqq \left|\int_a^x [f(x, \alpha + \Delta\alpha) - f(x, \alpha)]\,dx\right| + |R(x, \alpha + \Delta\alpha)| + |R(x, \alpha)|.$$

* It is of course assumed that $f(x, \alpha)$ is continuous in (x, α) for all values of x and α under consideration, and in the theorem on differentiation it is further assumed that $f'_\alpha(x, \alpha)$ is continuous.

† It should be noticed, however, that although the conditions which have been imposed are *sufficient* to establish the theorems, they are *not necessary*; that is, it may happen that the function will be continuous and that its derivative and integral may be obtained by operating under the sign although the convergence is not uniform. In this case a special investigation would have to be undertaken; and if no process for justifying the continuity, integration, or differentiation could be devised, it might be necessary in the case of an integral occurring in some application to assume that the formal work led to the right result if the result looked reasonable from the point of view of the problem under discussion, — the chance of getting an erroneous result would be tolerably small.

Now let x be taken so large that $|R|<\epsilon$ for all α's and for all larger values of x — the condition of uniformity. Then the finite integral (§ 118)

$$\int_a^x f(x, \alpha)\,dx \quad \text{is continuous in } \alpha \text{ and hence} \quad \int_a^x [f(x, \alpha+\Delta\alpha)-f(x, \alpha)]\,dx$$

can be made less than ϵ by taking $\Delta\alpha$ small enough. Hence $|\Delta\psi|<3\epsilon$; that is, by taking $\Delta\alpha$ small enough the quantity $|\Delta\psi|$ may be made less than any assigned number 3ϵ. The continuity is therefore proved.

To prove the integrability under the sign a like use is made of the condition of uniformity and of the earlier proof for a finite integral (§ 120).

$$\int_{\alpha_0}^{\alpha_1}\psi(\alpha)\,d\alpha=\int_{\alpha_0}^{\alpha_1}\int_a^x f(x, \alpha)\,dx\,d\alpha+\int_{\alpha_0}^{\alpha_1}R\,dx=\int_a^x\int_{\alpha_0}^{\alpha_1}f(x, \alpha)\,d\alpha\,dx+\zeta.$$

Now let x become infinite. The quantity ζ can approach no other limit than 0; for by taking x large enough $R<\epsilon$ and $|\zeta|<\epsilon(\alpha_1-\alpha_0)$ independently of α. Hence as x becomes infinite, the integral converges to the constant expression on the left and

$$\int_{\alpha_0}^{\alpha_1}\psi(\alpha)\,d\alpha=\int_a^\infty\int_{\alpha_0}^{\alpha_1}f(x, \alpha)\,d\alpha\,dx.$$

Moreover if the integration be to a variable limit for α, then

$$\Psi(\alpha)=\int_{\alpha_0}^{\alpha}\psi(\alpha)\,d\alpha=\int_a^\infty\int_{\alpha_0}^{\alpha}f(x, \alpha)\,d\alpha\,dx=\int_a^\infty F(x, \alpha)\,dx.$$

Also $$\left|\int_x^\infty F(x, \alpha)\,dx\right|=\left|\int_x^\infty\int_{\alpha_0}^{\alpha}f(x, \alpha)\,d\alpha\,dx\right|=\left|\int_{\alpha_0}^{\alpha}\int_x^\infty f(x, \alpha)\,dx\,d\alpha\right|<\epsilon(\alpha-\alpha_0).$$

Hence it appears that the remainder for the new integral is less than $\epsilon(\alpha_1-\alpha_0)$ for all values of α; the convergence is therefore uniform and a second integration may be performed if desired. Thus *if an infinite integral converges uniformly, it may be integrated as many times as desired under the sign.* It should be noticed that the proof fails to cover the case of integration to an infinite upper limit for α.

For the case of differentiation it is necessary to show that

$$\int_a^\infty f'_\alpha(x, \alpha_\xi)\,dx=\phi'(\alpha_\xi).\qquad \text{Consider } \int_a^\infty f'_\alpha(x, \alpha)\,dx=\omega(\alpha).$$

As the infinite integral is assumed to converge uniformly by the statement of the theorem, it is possible to integrate with respect to α under the sign. Then

$$\int_{\alpha_\xi}^{\alpha}\omega(\alpha)\,d\alpha=\int_a^\infty\int_{\alpha_\xi}^{\alpha}f'_\alpha(x, \alpha)\,d\alpha\,dx=\int_a^\infty[f(x, \alpha)-f(x, \alpha_\xi)]\,dx=\phi(\alpha)-\phi(\alpha_\xi).$$

The integral on the left may be differentiated with respect to α, and hence $\phi(\alpha)$ must be differentiable. The differentiation gives $\omega(\alpha)=\phi'(\alpha)$ and hence $\omega(\alpha_\xi)=\phi'(\alpha_\xi)$. The theorem is therefore proved. This theorem and the two above could be proved in analogous ways in the case of an infinite integral due to the fact that the integrand $f(x, \alpha)$ became infinite at the ends of (or within) the interval of integration with respect to x; the proofs need not be given here.

145. The method of integrating or differentiating under the sign of integration may be applied to evaluate infinite integrals when the conditions of uniformity are properly satisfied, in precisely the same manner as the method was previously applied to the case of finite integrals where

the question of the uniformity of convergence did not arise (§§ 119–120). The examples given below will serve to illustrate how the method works and in particular to show how readily the test for uniformity may be applied in some cases. Some of the examples are purposely chosen identical with some which have previously been treated by other methods.

Consider first an integral which may be found by direct integration, namely,

$$\int_0^\infty e^{-ax}\cos bx dx = \frac{a}{a^2+b^2}. \qquad \text{Compare } \int_0^\infty e^{-ax}dx = \frac{1}{a}.$$

The integrand e^{-ax} is a positive quantity greater than or equal to $e^{-ax}\cos bx$ for all values of b. Hence, by the general test, the first integral regarded as a function of b converges uniformly for all values of b, defines a continuous function, and may be integrated between any limits, say from 0 to b. Then

$$\int_0^b\int_0^\infty e^{-ax}\cos bx dx db = \int_0^\infty\int_0^b e^{-ax}\cos bx db dx$$

$$= \int_0^\infty e^{-ax}\frac{\sin bx}{x}dx = \int_0^b \frac{adb}{a^2+b^2} = \tan^{-1}\frac{b}{a}.$$

Integrate again. $\displaystyle\int_0^\infty\int_0^b e^{-ax}\frac{\sin bx}{x}db dx = \int_0^\infty e^{-ax}\frac{1-\cos bx}{x^2}dx$

$$= b\tan^{-1}\frac{b}{a} - \frac{a}{2}\log(a^2+b^2).$$

Compare $\displaystyle\int_0^\infty e^{-ax}\frac{1-\cos bx}{x^2}dx$ and $\displaystyle\int_0^\infty \frac{1-\cos bx}{x^2}dx.$

Now as the second integral has a positive integrand which is never less than the integrand of the first for any positive value of a, the first integral converges uniformly for all positive values of a including 0, is a continuous function of a, and the value of the integral for $a=0$ may be found by setting a equal to 0 in the integrand. Then

$$\int_0^\infty \frac{1-\cos bx}{x^2}dx = \lim_{a\doteq 0}\left[b\tan^{-1}\frac{b}{a} - \frac{a}{2}\log(a^2+b^2)\right] = |b|\frac{\pi}{2}.$$

The change of the variable to $x' = \frac{1}{2}x$ and an integration by parts give respectively

$$\int_0^\infty \frac{\sin^2 bx}{x^2}dx = \frac{\pi}{2}|b|, \qquad \int_0^\infty \frac{\sin bx}{x}dx = +\frac{\pi}{2} \quad\text{or}\quad -\frac{\pi}{2}, \quad\text{as}\quad b>0 \quad\text{or}\quad b<0.$$

This last result might be obtained *formally* by taking the limit

$$\lim_{a\doteq 0}\int_0^\infty e^{-ax}\frac{\sin bx}{x}dx = \int_0^\infty \frac{\sin bx}{x}dx = \tan^{-1}\frac{b}{0} = \pm\frac{\pi}{2}$$

after the first integration; but such a process would be unjustifiable without first showing that the integral was a continuous function of a for small positive values of a and for 0. In this case $|x^{-1}e^{-ax}\sin bx| \leqq |x^{-1}\sin x|$, but as the integral of $|x^{-1}\sin bx|$ does not converge, the test for uniformity fails to apply. Hence the limit would not be justified without special investigation. Here the limit does give the right result, but a simple case where the integral of the limit is not the limit of the integral is

$$\lim_{b\doteq 0}\int_0^\infty \frac{\sin bx}{x}dx = \lim_{b\doteq 0}\left(\pm\frac{\pi}{2}\right) = \pm\frac{\pi}{2} \neq \int_0^\infty \lim_{b\doteq 0}\frac{\sin bx}{x}dx = \int_0^\infty \frac{0}{x}dx = 0.$$

As a second example consider the evaluation of $\int_0^\infty e^{-\left(x-\frac{a}{x}\right)^2}dx$. Differentiate.

$$\phi'(a) = \frac{d}{da}\int_0^\infty e^{-\left(x-\frac{a}{x}\right)^2}dx - 2\int_0^\infty e^{-\left(x-\frac{a}{x}\right)^2}\left(x-\frac{a}{x}\right)\frac{1}{x}dx$$
$$= 2\int_0^\infty e^{-\left(x-\frac{a}{x}\right)^2}\left(1-\frac{a}{x^2}\right)dx.$$

To justify the differentiation this last integral must be shown to converge uniformly. In the first place note that the integrand does not become infinite at the origin, although one of its factors does. Hence the integral is infinite only by virtue of its infinite limit. Suppose $a \geqq 0$; then for large values of x

$$e^{-\left(x-\frac{a}{x}\right)^2}\left(1-\frac{a}{x^2}\right) \leqq e^{2a}e^{-x^2} \quad \text{and} \quad \int^\infty e^{-x^2}dx \quad \text{converges (§ 143).}$$

Hence the convergence is uniform when $a \geqq 0$, and the differentiation is justified. But, by the change of variable $x' = -a/x$, when $a > 0$,

$$\int_0^\infty e^{-\left(x-\frac{a}{x}\right)^2}\frac{adx}{x^2} = \int_0^\infty e^{-\left(-\frac{a}{x'}+x'\right)^2}dx' = \int_0^\infty e^{-\left(x-\frac{a}{x}\right)^2}dx.$$

Hence the derivative above found is zero; $\phi'(a) = 0$ and

$$\phi(a) = \int_0^\infty e^{-\left(x-\frac{a}{x}\right)^2}dx = \text{const.} = \int_0^\infty e^{-x^2}dx = \tfrac{1}{2}\sqrt{\pi};$$

for the integral converges uniformly when $a \geqq 0$ and its constant value may be obtained by setting $a = 0$. As the convergence is uniform for any range of values of a, the function is everywhere continuous and equal to $\frac{1}{2}\sqrt{\pi}$.

As a third example calculate the integral $\phi(b) = \int_0^\infty e^{-a^2x^2}\cos bx dx$. Now

$$\frac{d\phi}{db} = \int_0^\infty -xe^{-a^2x^2}\sin bx dx = \frac{1}{2a^2}\left[e^{-a^2x^2}\sin bx\right]_0^\infty - \frac{b}{2a^2}\int_0^\infty e^{-a^2x^2}\cos bx dx.$$

The second step is obtained by integration by parts. The previous differentiation is justified by the fact that the integral of $xe^{-a^2x^2}$, which is greater than the integrand of the derived integral, converges. The differential equation may be solved.

$$\frac{d\phi}{db} = -\frac{b}{2a^2}\phi, \qquad \phi = Ce^{-\frac{b^2}{4a^2}}, \qquad \phi(0) = \int_0^\infty e^{-a^2x^2}dx = \frac{\sqrt{\pi}}{2a}.$$

Hence $$\phi(b) = \phi(0)e^{-\frac{b^2}{4a^2}} = \int_0^\infty e^{-a^2x^2}\cos bx dx = \frac{\sqrt{\pi}e^{-\frac{b^2}{4a^2}}}{2a}.$$

In determining the constant C, the function $\phi(b)$ is assumed continuous, as the integral for $\phi(b)$ obviously converges uniformly for all values of b.

146. The question of the integration under the sign is naturally connected with the question of infinite double integrals. The double integral $\int f(x, y)dA$ over an area A is said to be an infinite integral if that area extends out indefinitely in any direction or if the function $f(x, y)$ becomes infinite at any point of the area. The definition of

convergence is analogous to that given before in the case of infinite simple integrals. If the area A is infinite, it is replaced by a finite area A' which is allowed to expand so as to cover more and more of the area A. If the function $f(x, y)$ becomes infinite at a point or along a line in the area A, the area A is replaced by an area A' from which the singularities of $f(x, y)$ are excluded, and again the area A' is allowed to expand and approach coincidence with A. If then the double integral extended over A' approaches a definite limit which is independent of how A' approaches A, the double integral is said to converge. As

$$\iint f(x, y)\, dx dy = \iint J\left(\frac{x, y}{u, v}\right) f(\phi, \psi)\, du dv,$$

where $x = \phi(u, v)$, $y = \psi(u, v)$, is the rule for the change of variable and is applicable to A', it is clear that if either side of the equality approaches a limit which is independent of how A' approaches A, the other side must approach the same limit.

The theory of infinite double integrals presents numerous difficulties, the solution of which is beyond the scope of this work. It will be sufficient to point out in a simple case the questions that arise, and then state without proof a theorem which covers the cases which arise in practice. Suppose the region of integration is a complete quadrant so that the limits for x and y are 0 and ∞. The first question is, If the double integral converges, may it be evaluated by successive integration as

$$\int f(x, y)\, dA = \int_{x=0}^{\infty} \int_{y=0}^{\infty} f(x, y)\, dy dx = \int_{y=0}^{\infty} \int_{x=0}^{\infty} f(x, y)\, dx dy?$$

And conversely, if one of the iterated integrals converges so that it may be evaluated, does the other one, and does the double integral, converge to the same value? A part of this question also arises in the case of a function defined by an infinite integral. For let

$$\phi(x) = \int_{y=0}^{\infty} f(x, y)\, dy \quad \text{and} \quad \int_{x=0}^{\infty} \phi(x)\, dx = \int_{x=0}^{\infty} \int_{y=0}^{\infty} f(x, y)\, dy dx,$$

it being assumed that $\phi(x)$ converges except possibly for certain values of x, and that the integral of $\phi(x)$ from 0 to ∞ converges. The question arises, May the integral of $\phi(x)$ be evaluated by integration under the sign? The proofs given in § 144 for uniformly convergent integrals integrated over a finite region do not apply to this case of an infinite integral. In any particular given integral special methods may possibly be devised to justify for that case the desired transformations. But most cases are covered by a theorem due to de la Vallée Poussin: *If the*

function $f(x, y)$ does not change sign and is continuous except over a finite number of lines parallel to the axes of x and y, then the three integrals

$$\int f(x, y)\,dA, \quad \int_{x=0}^{\infty}\int_{y=0}^{\infty} f(x, y)\,dy\,dx, \quad \int_{y=0}^{\infty}\int_{x=0}^{\infty} f(x, y)\,dx\,dy, \quad (12)$$

*cannot lead to different determinate results; that is, if any two of them lead to definite results, those results are equal.** The chief use of the theorem is to establish the equality of the two iterated integrals when each is known to converge; the application requires no test for uniformity and is very simple.

As an example of the use of the theorem consider the evaluation of

$$I = \int_0^\infty e^{-x^2}dx = \int_0^\infty \alpha e^{-\alpha^2 x^2}dx.$$

Multiply by $e^{-\alpha^2}$ and integrate from 0 to ∞ with respect to α.

$$Ie^{-\alpha^2} = \int_0^\infty \alpha e^{-\alpha^2(1+x^2)}dx, \qquad I\int_0^\infty e^{-\alpha^2}d\alpha = I^2 = \int_0^\infty\int_0^\infty \alpha e^{-\alpha^2(1+x^2)}\,dx\,d\alpha.$$

Now the integrand of the iterated integral is positive and the integral, being equal to I^2, has a definite value. If the order of integrations is changed, the integral

$$\int_0^\infty\int_0^\infty \alpha e^{-\alpha^2(1+x^2)}\,d\alpha\,dx = \int_0^\infty \frac{1}{1+x^2}\frac{dx}{2} = \frac{1}{2}\tan^{-1}\infty = \frac{\pi}{4}$$

is seen also to lead to a definite value. Hence the values I^2 and $\frac{1}{4}\pi$ are equal.

EXERCISES

1. Note that the two integrands are continuous functions of (x, α) in the whole region $0 \leqq \alpha < \infty$, $0 \leqq x < \infty$ and that for each value of α the integrals converge. Establish the forms given to the remainders and from them show that it is not possible to take x so large that for all values of α the relation $|R(x, \alpha)| < \epsilon$ is satisfied, but may be satisfied for all α's such that $0 < \alpha_0 \leqq \alpha$. Hence infer that the convergence is nonuniform about $\alpha = 0$, but uniform elsewhere. Note that the functions defined are not continuous at $\alpha = 0$, but are continuous for all other values.

$$(\alpha) \int_0^\infty \alpha e^{-\alpha x}dx, \quad R(x, \alpha) = \int_x^\infty \alpha e^{-\alpha x}dx = e^{-\alpha x} - 1.$$

$$(\beta) \int_0^\infty \frac{\sin \alpha x}{x}dx, \quad R(x, \alpha) = \int_x^\infty \frac{\sin \alpha x}{x}dx = \int_{\alpha x}^\infty \frac{\sin x}{x}dx.$$

2. Repeat in detail the proofs relative to continuity, integration, and differentiation in case the integral is infinite owing to an infinite integrand at $x = b$.

* The theorem may be generalized by allowing $f(x, y)$ to be discontinuous over a finite number of curves each of which is cut in only a finite limited number of points by lines parallel to the axis. Moreover, the function may clearly be allowed to change sign to a certain extent, as in the case where $f > 0$ when $x > a$, and $f < 0$ when $0 < x < a$, etc., where the integral over the whole region may be resolved into the sum of a finite number of integrals. Finally, if the integrals are absolutely convergent and the integrals of $f(x, y)$ lead to definite results, so will the integrals of $f(x, y)$.

3. Show that differentiation under the sign is allowable in the following cases, and hence derive the results that are given:

(α) $\int_0^\infty e^{-\alpha x^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}$, $\alpha > 0$, $\int_0^\infty x^{2n}e^{-\alpha x^2}dx = \frac{\sqrt{\pi}}{2}\frac{1\cdot 3\cdots(2n-1)}{2^n\alpha^{n+\frac{1}{2}}}$,

(β) $\int_0^\infty xe^{-\alpha x^2}dx = \frac{1}{2\alpha}$, $\alpha > 0$. $\int_0^\infty x^{2n+1}e^{-\alpha x^2}dx = \frac{1\cdot 2\cdots n}{2\,\alpha^{n+1}}$,

(γ) $\int_0^\infty \frac{dx}{x^2+k} = \frac{\pi}{2}\frac{1}{\sqrt{k}}$, $k > 0$, $\int_0^\infty \frac{dx}{(x^2+k)^{n+1}} = \frac{\pi}{2}\frac{1\cdot 3\cdots(2n-1)}{2^n n!\,k^{n+\frac{1}{2}}}$,

(δ) $\int_0^1 x^n dx = \frac{1}{n+1}$, $n > -1$, $\int_0^1 x^n(-\log x)^m dx = \frac{m!}{(n+1)^{m+1}}$,

(ϵ) $\int_0^\infty \frac{x^{\alpha-1}}{1+x}dx = \frac{\pi}{\sin\alpha\pi}$. $0 < \alpha < 1$. $\int_0^\infty \frac{x^{\alpha-1}\log x}{1+x}dx = \frac{\pi^2\cos\alpha\pi}{\cos^2\alpha\pi - 1}$.

4. Establish the right to integrate and hence evaluate these:

(α) $\int_0^\infty e^{-\alpha x}dx$, $0 < \alpha_0 \leqq \alpha$. $\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}dx = \log\frac{b}{a}$, $b, a \geqq \alpha_0$.

(β) $\int_0^1 x^\alpha dx$. $-1 < \alpha_0 < \alpha$. $\int_0^1 \frac{x^a - x^b}{\log x}dx = \log\frac{a+1}{b+1}$, $b, a \geqq \alpha_0$.

(γ) $\int_0^\infty e^{-\alpha x}\cos mx\,dx$, $0 < \alpha_0 \leqq \alpha$. $\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}\cos mx\,dx = \frac{1}{2}\log\frac{b^2+m^2}{a^2+m^2}$.

(δ) $\int_0^\infty e^{-\alpha x}\sin mx\,dx$. $0 < \alpha_0 \leqq \alpha$. $\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}\sin mx\,dx = \tan^{-1}\frac{b}{m} - \tan^{-1}\frac{a}{m}$.

(ϵ) $\int_0^\infty e^{-\alpha^2x^2}dx = \frac{\sqrt{\pi}}{2\alpha}$, $0 < \alpha_0 \leqq \alpha$. $\int_0^\infty e^{-\frac{a^2}{x^2}} - e^{-\frac{b^2}{x^2}}dx = (b-a)\sqrt{\pi}$.

5. Evaluate: (α) $\int_0^\infty e^{-\alpha x}\frac{\sin\beta x}{x}dx = \tan^{-1}\frac{\beta}{\alpha}$,

(β) $\int_0^\infty e^{-x}\frac{1-\cos\alpha x}{x}dx = \log\sqrt{1+\alpha^2}$. (γ) $\int_0^\infty e^{-x^2}\frac{\sin 2\alpha x}{x}dx$,

(δ) $\int_0^\infty e^{-\left(x^2+\frac{a^2}{x^2}\right)}dx = \frac{\sqrt{\pi}}{2}e^{-2a}$, $a \geqq 0$. (ϵ) $\int_0^\infty \frac{\log(1+a^2x^2)}{1+b^2x^2}dx$.

6. If $0 < a < b$, obtain from $\int_0^\infty e^{-rx^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{r}}$ and justify the relations:

$$\int_a^b \frac{\sin r}{\sqrt{r}}dr = \frac{2}{\sqrt{\pi}}\int_a^b\int_0^\infty e^{-rx^2}\sin r\,dx\,dr = \frac{2}{\sqrt{\pi}}\int_0^\infty\int_a^b e^{-rx^2}\sin r\,dr\,dx$$

$$= \frac{2}{\sqrt{\pi}}\left[\sin a\int_0^\infty \frac{e^{-ax^2}x^2dx}{1+x^4} - \sin b\int_0^\infty \frac{e^{-bx^2}x^2dx}{1+x^4}\right.$$

$$\left.+\cos a\int_0^\infty \frac{e^{-ax^2}dx}{1+x^4} - \cos b\int_0^\infty \frac{e^{-bx^2}dx}{1+x^4}\right].$$

$$\int_0^r \frac{\sin r}{\sqrt{r}}dr = \sqrt{\frac{\pi}{2}} - \frac{2}{\sqrt{\pi}}\left[\sin r\int_0^\infty \frac{e^{-rx^2}x^2dx}{1+x^4} + \cos r\int_0^\infty \frac{e^{-rx^2}dx}{1+x^4}\right].$$

Similarly, $\displaystyle\int_0^r \frac{\cos r}{\sqrt{r}}dr = \sqrt{\frac{\pi}{2}} - \frac{2}{\pi}\left[\cos r\int_0^\infty \frac{e^{-rx^2}x^2dx}{1+x^4} - \sin r\int_0^\infty \frac{e^{-rx^2}dx}{1+x^4}\right].$

Also $\displaystyle\int_0^\infty \frac{\sin r}{\sqrt{r}}dr = \int_0^\infty \frac{\cos r}{\sqrt{r}}dr = \sqrt{\frac{\pi}{2}},\quad \int_0^\infty \sin\frac{\pi}{2}r^2dr = \int_0^\infty \cos\frac{\pi}{2}r^2dr = \frac{1}{2}.$

7. Given that $\displaystyle\frac{1}{1+x^2} = 2\int_0^\infty \alpha e^{-\alpha^2(1+x^2)}d\alpha$, show that

$$\int_0^\infty \frac{1+\cos mx}{1+x^2}dx = \frac{\pi}{2}(1+e^{-m}) \quad\text{and}\quad \int_0^\infty \frac{\cos mx}{1+x^2}dx = \frac{\pi}{2}e^{-m},\quad m>0.$$

8. Express $R(x, \alpha) = \displaystyle\int_x^\infty \frac{x\sin\alpha x}{1+x^2}dx$, by integration by parts and also by substituting x' for αx, in such a form that the uniform convergence for α such that $0<\alpha_0 \leqq \alpha$ is shown. Hence from Ex. 7 prove

$$\int_0^\infty \frac{x\sin\alpha x}{1+x^2}dx = \frac{\pi}{2}e^{-\alpha},\qquad \alpha>0 \qquad \text{(by differentiation).}$$

Show that this integral does not satisfy the test for uniformity given in the text; also that for $\alpha=0$ the convergence is not uniform and that the integral is also discontinuous.

9. If $f(x, \alpha, \beta)$ is continuous in (x, α, β) for $0\leqq x<\infty$ and for all points (α, β) of a region in the $\alpha\beta$-plane, and if the integral $\phi(\alpha, \beta) = \displaystyle\int_0^\infty f(x, \alpha, \beta)dx$ converges uniformly for said values of (α, β), show that $\phi(\alpha, \beta)$ is continuous in (α, β). Show further that if $f'_\alpha(x, \alpha, \beta)$ and $f'_\beta(x, \alpha, \beta)$ are continuous and their integrals converge uniformly for said values of (α, β), then

$$\int_0^\infty f'_\alpha(x, \alpha, \beta)dx = \phi'_\alpha,\qquad \int_0^\infty f'_\beta(x, \alpha, \beta)dx = \phi'_\beta,$$

and ϕ'_α, ϕ'_β are continuous in (α, β). The proof in the text holds almost verbatim.

10. If $f(x, \gamma) = f(x, \alpha+i\beta)$ is a function of x and the complex variable $\gamma = \alpha + i\beta$ which is continuous in (x, α, β), that is, in (x, γ) over a region of the γ-plane, etc., as in Ex. 9, and if $f'_\gamma(x, \gamma)$ satisfies the same conditions, show that

$$\phi(\gamma) = \int_0^\infty f(x, \gamma)dx \text{ defines an analytic function of } \gamma \text{ in said region.}$$

11. Show that $\displaystyle\int_0^\infty e^{-\gamma x^2}dx$, $\gamma = \alpha+i\beta$, $\alpha \geqq \alpha_0 > 0$, defines an analytic function of γ over the whole γ-plane to the right of the vertical $\alpha=\alpha_0$. Hence infer

$$\phi(\gamma) = \int_0^\infty e^{-\gamma x^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{\gamma}} = \frac{1}{2}\sqrt{\frac{\pi}{\alpha+i\beta}},\qquad \alpha\geqq\alpha_0>0.$$

Prove
$$\int_0^\infty e^{-\alpha x^2}\cos\beta x^2dx = \frac{1}{2}\sqrt{\frac{\pi}{2}\frac{\alpha+\sqrt{\alpha^2+\beta^2}}{\alpha^2+\beta^2}},$$

$$\int_0^\infty e^{-\alpha x^2}\sin\beta x^2dx = \frac{1}{2}\sqrt{\frac{\pi}{2}\frac{-\alpha+\sqrt{\alpha^2+\beta^2}}{\alpha^2+\beta^2}}.$$

12. Integrate $\int_x^\infty \frac{1}{x} e^{-\alpha x^2} x \cos \beta x^2 dx$ of Ex. 11 by parts with $x \cos \beta x^2 dx = du$ to show that the convergence is uniform at $\alpha = 0$. Hence find $\int_0^\infty \cos \beta x^2 dx$.

13. From $\int_{-\infty}^{+\infty} \cos x^2 dx = \int_{-\infty}^{+\infty} \cos (x + \alpha)^2 dx = \sqrt{\frac{\pi}{2}} = \int_{-\infty}^{+\infty} \sin (x + \alpha)^2 dx$, with the results $\int_{-\infty}^{+\infty} \cos x^2 \sin 2\alpha x dx = \int_{-\infty}^{+\infty} \sin x^2 \sin 2\alpha x dx = 0$ due to the fact that $\sin x$ is an odd function, establish the relations

$$\int_0^\infty \cos x^2 \cos 2\alpha x dx = \frac{\sqrt{\pi}}{2} \cos \left(\frac{\pi}{4} - \alpha^2\right), \quad \int_0^\infty \sin x^2 \cos 2\alpha x dx = \frac{\sqrt{\pi}}{2} \sin \left(\frac{\pi}{4} - \alpha^2\right).$$

14. Calculate: $(\alpha)\ \int_0^\infty e^{-a^2x^2} \cosh bx dx$, $\quad (\beta)\ \int_0^\infty x e^{-ax} \cos bx dx$,

and (together) $\quad (\gamma)\ \int_0^\infty \cos \left(\frac{x^2}{2} \pm \frac{\alpha^2}{2x^2}\right) dx$, $\quad (\delta)\ \int_0^\infty \sin \left(\frac{x^2}{2} \pm \frac{\alpha^2}{2x^2}\right) dx$.

15. In continuation of Exs. 10–11, p. 368, prove at least formally the relations:

$$\lim_{k=\infty} \int_a^0 f(x) \frac{\sin kx}{x} dx = \frac{\pi}{2} f(0), \qquad \lim_{k=\infty} \frac{1}{\pi} \int_{-a}^a f(x) \frac{\sin kx}{x} dx = f(0),$$

$$\int_0^k \int_{-a}^a f(x) \cos kx dx dk = \int_{-a}^a \int_0^k f(x) \cos kx dk dx = \int_{-a}^a f(x) \frac{\sin kx}{x} dx,$$

$$\frac{1}{\pi} \int_0^\infty \int_{-a}^a f(x) \cos kx dx dk = \lim_{k=\infty} \frac{1}{\pi} \int_{-a}^a f(x) \frac{\sin kx}{x} dx = f(0),$$

$$\frac{1}{\pi} \int_0^\infty \int_{-\infty}^\infty f(x) \cos kx dx dk = f(0), \qquad \frac{1}{\pi} \int_0^\infty \int_{-\infty}^\infty f(x) \cos k (x - t) dx dk = f(t).$$

The last form is known as Fourier's Integral; it represents a function $f(t)$ as a double infinite integral containing a parameter. Wherever possible, justify the steps after placing sufficient restrictions on $f(x)$.

16. From $\int_0^\infty e^{-xy} dy = \frac{1}{x}$ prove $\int_0^\infty \frac{e^{-ax} - e^{-bx}}{x} dx = \log \frac{b}{a}\cdot$ Prove also

$$\int_0^\infty x^{n-1} e^{-x} dx \int_0^\infty x^{m-1} e^{-x} dx$$
$$= 2 \int_0^\infty r^{2n+2m-2} e^{-r^2} dr^2 \int_0^{\frac{\pi}{2}} \sin^{2n-1}\phi \cos^{2m-1}\phi d\phi.$$

17. Treat the integrals (12) by polar coördinates and show that

$$\int f(x, y) dA = \int_0^{\frac{\pi}{2}} \int_0^\infty f(r \cos \phi, r \sin \phi) r dr d\phi$$

will converge if $|f| < r^{-2-k}$ as r becomes infinite. If $f(x, y)$ becomes infinite at the origin, but $|f| < r^{-2+k}$, the integral converges as r approaches zero. Generalize these results to triple integrals and polar coördinates in space; the only difference is that 2 becomes 3.

18. As in Exs. 1, 8, 12, uniformity of convergence may often be tested directly, without the test of page 369; treat the integrand $x^{-1} e^{-ax} \sin bx$ of page 371, where that test failed.

CHAPTER XIV

SPECIAL FUNCTIONS DEFINED BY INTEGRALS

147. The Gamma and Beta functions. The two integrals

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx, \qquad B(m, n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx \tag{1}$$

converge when $n > 0$ and $m > 0$, and hence define functions of the parameters n or n and m for all positive values, zero not included. Other forms may be obtained by changes of variable. Thus

$$\Gamma(n) = 2\int_0^\infty y^{2n-1}e^{-y^2}dy, \qquad \text{by} \quad x = y^2, \tag{2}$$

$$\Gamma(n) = \int_0^1 \left(\log\frac{1}{y}\right)^{n-1} dy, \qquad \text{by} \quad e^{-x} = y, \tag{3}$$

$$B(m, n) = \int_0^1 y^{n-1}(1-y)^{m-1}dy = B(n, m), \quad \text{by} \quad x = 1 - y, \tag{4}$$

$$B(m, n) = \int_0^\infty \frac{y^{m-1}dy}{(1+y)^{m+n}}, \qquad \text{by} \quad x = \frac{y}{1+y}, \tag{5}$$

$$B(m, n) = 2\int_0^{\frac{\pi}{2}} \sin^{2m-1}\phi \cos^{2n-1}\phi d\phi, \qquad \text{by} \quad x = \sin^2\phi. \tag{6}$$

If the original form of $\Gamma(n)$ be integrated by parts, then

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx = \frac{1}{n}x^n e^{-x}\Big]_0^\infty + \frac{1}{n}\int_0^\infty x^n e^{-x}dx = \frac{1}{n}\Gamma(n+1).$$

The resulting relation $\Gamma(n+1) = n\Gamma(n)$ shows that the values of the Γ-function for $n+1$ may be obtained from those for n, and that consequently the values of the function will all be determined if the values over a unit interval are known. Furthermore

$$\begin{aligned} \Gamma(n+1) &= n\Gamma(n) = n(n-1)\Gamma(n-1) \\ &= n(n-1)\cdots(n-k)\Gamma(n-k) \end{aligned} \tag{7}$$

is found by successive reduction, where k is any integer less than n. If in particular n is an integer and $k = n - 1$, then

$$\Gamma(n+1) = n(n-1)\cdots 2\cdot 1\cdot\Gamma(1) = n!\,\Gamma(1) = n!; \tag{8}$$

since when $n=1$ a direct integration shows that $\Gamma(1)=1$. Thus *for integral values of n the Γ-function is the factorial;* and for other than integral values it may be regarded as a sort of generalization of the factorial.

Both the Γ- and B-functions are continuous for all values of the parameters greater than, but not including, zero. To prove this it is sufficient to show that the convergence is uniform. Let n be any value in the interval $0 < n_0 \leqq n \leqq N$; then

$$\int_0 x^{n-1}e^{-x}dx \leqq \int_0 x^{n_0-1}e^{-x}dx, \qquad \int^\infty x^{n-1}e^{-x}dx \leqq \int^\infty x^{N-1}e^{-x}dx.$$

The two integrals converge and the general test for uniformity (§ 144) therefore applies; the application at the lower limit is not necessary except when $n<1$. Similar tests apply to $B(m, n)$. Integration with respect to the parameter may therefore be carried under the sign. The derivatives

$$\frac{d^k\Gamma(n)}{dn^k} = \int_0^\infty x^{n-1}e^{-x}(\log x)^k dx \tag{9}$$

may also be had by differentiating under the sign; for these derived integrals may likewise be shown to converge uniformly.

By multiplying two Γ-functions expressed as in (2), treating the product as an iterated or double integral extended over a whole quadrant, and evaluating by transformation to polar coördinates (all of which is justifiable by § 146, since the integrands are positive and the processes lead to a determinate result), the B-function may be expressed in terms of the Γ-function.

$$\Gamma(n)\Gamma(m) = 4\int_0^\infty x^{2n-1}e^{-x^2}dx\int_0^\infty y^{2m-1}e^{-y^2}dy = 4\int_0^\infty\int_0^\infty x^{2n-1}y^{2m-1}e^{-x^2-y^2}dxdy$$

$$= 4\int_0^\infty r^{2n+2m-1}e^{-r^2}dr\int_0^{\frac{\pi}{2}}\sin^{2m-1}\phi\cos^{2n-1}\phi d\phi = \Gamma(n+m)B(m, n).$$

Hence
$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = B(n, m). \tag{10}$$

The result is symmetric in m and n, as must be the case inasmuch as the B-function has been seen by (4) to be symmetric.

That $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ follows from (9) of § 143 after setting $n=\frac{1}{2}$ in (2); it may also be deduced from a relation of importance which is obtained from (10) and (5), and from (8) of § 142, namely, if $n<1$,

$$\frac{\Gamma(n)\Gamma(1-n)}{\Gamma(1)=1} = B(n, 1-n) = \int_0^\infty \frac{y^{n-1}}{1+y}dy = \frac{\pi}{\sin n\pi},$$

or
$$\Gamma(n)\Gamma(1-n) = \frac{\pi}{\sin n\pi}. \tag{11}$$

As it was seen that all values of $\Gamma(n)$ could be had from those in a unit interval, say from 0 to 1, the relation (11) shows that the interval may be further reduced to $\frac{1}{2} \leqq n \leqq 1$ and that the values for the interval $0 < 1 - n < \frac{1}{2}$ may then be found.

148. By suitable changes of variable a great many integrals may be reduced to B- and Γ-integrals and thus expressed in terms of Γ-functions. Many of these types are given in the exercises below; a few of the most important ones will be taken up here. By $y = ax$,

$$\int_0^a x^{m-1}(a-x)^{n-1}dx = a^{m+n-1}\int_0^1 y^{m-1}(1-y)^{n-1}dy = a^{m+n-1}\mathrm{B}(m, n)$$

or
$$\int_0^a x^{m-1}(a-x)^{n-1}dx = a^{m+n-1}\frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}; \qquad a > 0. \tag{12}$$

Next let it be required to evaluate the triple integral

$$I = \iiint x^{l-1}y^{m-1}z^{n-1}dxdydz, \qquad x + y + z \leqq 1,$$

over the volume bounded by the coördinate planes and $x + y + z = 1$, that is, over all positive values of x, y, z such that $x + y + z \leqq 1$. Then

$$I = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} x^{l-1}y^{m-1}z^{n-1}dzdydx$$

$$= \frac{1}{n}\int_0^1 \int_0^{1-x} x^{l-1}y^{m-1}(1-x-y)^n dydx.$$

By (12) $\displaystyle\int_0^{1-x} y^{m-1}(1-x-y)^n dy = \frac{\Gamma(m)\Gamma(n+1)}{\Gamma(m+n+1)}(1-x)^{m+n}.$

Then
$$I = \frac{\Gamma(m)\Gamma(n+1)}{n\Gamma(m+n+1)}\int_0^1 x^{l-1}(1-x)^{m+n}dx$$

$$= \frac{\Gamma(m)\Gamma(n+1)}{n\Gamma(m+n+1)}\frac{\Gamma(l)\Gamma(m+n+1)}{\Gamma(l+m+n+1)}.$$

This result may be simplified by (7) and by cancellation. Then

$$I = \iiint x^{l-1}y^{m-1}z^{n-1}dxdydz = \frac{\Gamma(l)\Gamma(m)\Gamma(n)}{\Gamma(l+m+n+1)}. \tag{13}$$

There are simple modifications and generalizations of these results which are sometimes useful. For instance if it were desired to evaluate I over the range of positive values such that $x/a + y/b + z/c \leqq h$, the change $x = ah\xi$, $y = bh\eta$, $z = ch\zeta$ gives

$$I = a^l b^m c^n h^{l+m+n}\iiint \xi^{l-1}\eta^{m-1}\zeta^{n-1}d\xi d\eta d\zeta, \qquad \xi + \eta + \zeta \leqq 1.$$

$$I = \iiint x^{l-1}y^{m-1}z^{n-1}dxdydz = a^l b^m c^n\frac{\Gamma(l)\Gamma(m)\Gamma(n)}{\Gamma(l+m+n+1)}h^{l+m+n}, \qquad \frac{x}{a} + \frac{y}{b} + \frac{z}{c} \leqq h.$$

The value of this integral extended over the lamina between two parallel planes determined by the values h and $h + dh$ for the constant h would be

$$dI = a^l b^m c^n \frac{\Gamma(l)\Gamma(m)\Gamma(n)}{\Gamma(l+m+n)} h^{l+m+n-1}dh.$$

Hence if the integrand contained a function $f(h)$, the reduction would be

$$\iiint x^{l-1}y^{m-1}z^{n-1}f\left(\frac{x}{a}+\frac{y}{b}+\frac{z}{c}\right)dxdydz$$
$$= a^l b^m c^n \frac{\Gamma(l)\Gamma(m)\Gamma(n)}{\Gamma(l+m+n)}\int_0^H f(h)\,h^{l+m+n-1}dh$$

if the integration be extended over all values $x/a + y/b + z/c \leqq H$.

Another modification is to the case of the integral extended over a volume

$$J = \iiint x^{l-1}y^{m-1}z^{n-1}dxdydz, \qquad \left(\frac{x}{a}\right)^p + \left(\frac{y}{b}\right)^q + \left(\frac{z}{c}\right)^r \leqq h,$$

which is the octant of the surface $(x/a)^p + (y/b)^q + (z/c)^r = h$. The reduction to

$$J = \frac{a^l b^m c^n h^{\frac{l}{p}+\frac{m}{q}+\frac{n}{r}}}{pqr}\iiint \xi^{\frac{l}{p}-1}\eta^{\frac{m}{q}-1}\zeta^{\frac{n}{r}-1}d\xi d\eta d\zeta, \qquad \xi+\eta+\zeta \leqq 1,$$

is made by $\xi h = \left(\frac{x}{a}\right)^p$, $\eta h = \left(\frac{y}{b}\right)^q$, $\zeta h = \left(\frac{z}{c}\right)^r$, $dx = \frac{a}{p}h^{\frac{1}{p}}\xi^{\frac{1}{p}-1}, \cdots$.

$$J = \iiint x^{l-1}y^{m-1}z^{n-1}dxdydz = \frac{a^l b^m c^n}{pqr}\frac{\Gamma\left(\frac{l}{p}\right)\Gamma\left(\frac{m}{q}\right)\Gamma\left(\frac{n}{r}\right)}{\Gamma\left(\frac{l}{p}+\frac{m}{q}+\frac{n}{r}+1\right)}h^{\frac{l}{p}+\frac{m}{q}+\frac{n}{r}}.$$

This integral is of importance because the bounding surface here occurring is of a type tolerably familiar and frequently arising; it includes the ellipsoid, the surface $x^{\frac{1}{2}} + y^{\frac{1}{2}} + z^{\frac{1}{2}} = a^{\frac{1}{2}}$, the surface $x^{\frac{2}{3}} + y^{\frac{2}{3}} + z^{\frac{2}{3}} = a^{\frac{2}{3}}$. By taking $l = m = n = 1$ the volumes of the octants are expressed in terms of the Γ-function; by taking first $l = 3$, $m = n = 1$, and then $m = 3$, $l = n = 1$, and adding the results, the moments of inertia about the z-axis are found.

Although the case of a triple integral has been treated, the results for a double integral or a quadruple integral or integral of higher multiplicity are made obvious. For example,

$$\iint x^{l-1}y^{m-1}dxdy = a^l b^m h^{l+m}\frac{\Gamma(l)\Gamma(m)}{\Gamma(l+m+1)}, \qquad \frac{x}{a}+\frac{y}{b} \leqq h,$$

$$\iint x^{l-1}y^{m-1}dxdy = \frac{a^l b^m}{pq}\frac{\Gamma\left(\frac{l}{p}\right)\Gamma\left(\frac{m}{q}\right)}{\Gamma\left(\frac{l}{p}+\frac{m}{q}+1\right)}h^{\frac{l}{p}+\frac{m}{q}}, \qquad \left(\frac{x}{a}\right)^p + \left(\frac{y}{b}\right)^q \leqq h.$$

$$\iint x^{l-1}y^{m-1}f\left[\left(\frac{x}{a}\right)^p + \left(\frac{y}{b}\right)^q\right]dxdy = \frac{a^l b^m}{pq}\frac{\Gamma\left(\frac{l}{p}\right)\Gamma\left(\frac{m}{q}\right)}{\Gamma\left(\frac{l}{p}+\frac{m}{q}\right)}\int_0^H f(h)\,h^{\frac{l}{p}+\frac{m}{q}-1}dh,$$

$$\left(\frac{x}{a}\right)^p + \left(\frac{y}{b}\right)^q \leqq H,$$

$$\iiiint x^{k-1}y^{l-1}z^{m-1}t^{n-1}dxdydzdt = \frac{a^k b^l c^m d^n}{pqrs}\frac{\Gamma\left(\frac{k}{p}\right)\Gamma\left(\frac{l}{q}\right)\Gamma\left(\frac{m}{r}\right)\Gamma\left(\frac{n}{s}\right)}{\Gamma\left(\frac{k}{p}+\frac{l}{q}+\frac{m}{r}+\frac{n}{s}+1\right)},$$

$$\left(\frac{x}{a}\right)^p + \left(\frac{y}{b}\right)^q + \left(\frac{z}{c}\right)^r + \left(\frac{t}{d}\right)^s \leqq 1.$$

149. If the product (11) be formed for each of $\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}$, and the results be multiplied and reduced by Ex. 19 below, then

$$\Gamma\left(\frac{1}{n}\right)\Gamma\left(\frac{2}{n}\right)\cdots\Gamma\left(\frac{n-1}{n}\right) = \frac{(2\pi)^{\frac{n-1}{2}}}{\sqrt{n}}. \tag{14}$$

The logarithms may be taken and the result be divided by n.

$$\sum_{k=1}^{n} \log\Gamma\left(\frac{k}{n}\right)\cdot\frac{1}{n} = \left(\frac{1}{2}-\frac{1}{2n}\right)\log 2\pi - \frac{1}{2}\frac{\log n}{n}.$$

Now if n be allowed to become infinite, the sum on the left is that formed in computing an integral if $dx = 1/n$. Hence

$$\lim_{n=\infty}\sum \log\Gamma(x_i)\,\Delta x = \int_0^1 \log\Gamma(x)\,dx = \log\sqrt{2\pi}. \tag{15}$$

Then

$$\int_0^1 \log\Gamma(a+x)\,dx = a(\log a - 1) + \log\sqrt{2\pi} \tag{15'}$$

may be evaluated by differentiating under the sign (Ex. 12 (θ), p. 288).

By the use of differentiation and integration under the sign, the expressions for the first and second logarithmic derivatives of $\Gamma(n)$ and for $\log\Gamma(n)$ itself may be found as definite integrals. By (9) and the expression of Ex. 4 (α), p. 375, for $\log x$,

$$\Gamma'(n) = \int_0^\infty x^{n-1}e^{-x}\log x\,dx = \int_0^\infty x^{n-1}e^{-x}\int_0^\infty \frac{e^{-\alpha}-e^{-\alpha x}}{\alpha}\,d\alpha dx.$$

If the iterated integral be regarded as a double integral, the order of the integrations may be inverted; for the integrand maintains a positive sign in the region $1 < x < \infty$, $0 < \alpha < \infty$, and a negative sign in the region $0 < x < 1$, $0 < \alpha < \infty$, and the integral from 0 to ∞ in x may be considered as the sum of the integrals from 0 to 1 and from 1 to ∞, — to each of which the inversion is applicable (§ 146) because the integrand does not change sign and the results (to be obtained) are definite. Then by Ex. 1 (α),

$$\Gamma'(n) = \int_0^\infty\int_0^\infty x^{n-1}e^{-x}\frac{e^{-\alpha}-e^{-\alpha x}}{\alpha}\,dx d\alpha = \Gamma(n)\int_0^\infty\left(e^{-\alpha} - \frac{1}{(1+\alpha)^n}\right)\frac{d\alpha}{\alpha}$$

or $$\frac{\Gamma'(n)}{\Gamma(n)} = \frac{d}{dn}\log\Gamma(n) = \int_0^\infty \left(e^{-\alpha} - \frac{1}{(1+\alpha)^n}\right)\frac{d\alpha}{\alpha}. \tag{16}$$

This value may be simplified by subtracting from it the particular value $-\gamma = \Gamma'(1)/\Gamma(1) = \Gamma'(1)$ found for $n = 1$. Then

$$\frac{\Gamma'(n)}{\Gamma(n)} - \frac{\Gamma'(1)}{\Gamma(1)} = \frac{\Gamma'(n)}{\Gamma(n)} + \gamma = \int_0^\infty \left(\frac{1}{1+\alpha} - \frac{1}{(1+\alpha)^n}\right)\frac{d\alpha}{\alpha}.$$

The change of $1+\alpha$ to $1/\alpha$ or to e^α gives

$$\frac{\Gamma'(n)}{\Gamma(n)} + \gamma = \int_0^1 \frac{1-\alpha^{n-1}}{1-\alpha}d\alpha = \int_0^\infty \frac{e^{-\alpha} - e^{-\alpha n}}{1-e^{-\alpha}}d\alpha. \tag{17}$$

Differentiate: $$\frac{d^2}{dn^2}\log\Gamma(n) = \int_0^\infty \frac{\alpha e^{-\alpha n}}{1-e^{-\alpha}}d\alpha. \tag{18}$$

To find $\log\Gamma(n)$ integrate (16) from $n = 1$ to $n = n$. Then

$$\log\Gamma(n) = \int_0^\infty \left[(n-1)e^{-\alpha} - \frac{(1+\alpha)^{-1} - (1+\alpha)^{-n}}{\log(1+\alpha)}\right]\frac{d\alpha}{\alpha}, \tag{19}$$

since $\Gamma(1) = 1$ and $\log\Gamma(1) = 0$. As $\Gamma(2) = 1$,

$$\log\Gamma(2) = 0 = \int_0^\infty \left[\frac{e^{-\alpha}}{\alpha} - \frac{(1+\alpha)^{-2}}{\log(1+\alpha)}\right]d\alpha,$$

and $$\log\Gamma(n) = \int_0^\infty \left[\frac{n-1}{(1+\alpha)^2} - \frac{(1+\alpha)^{-1} - (1+\alpha)^{-n}}{\alpha}\right]\frac{d\alpha}{\log(1+\alpha)}$$

by subtracting from (19) the quantity $(n-1)\log\Gamma(2) = 0$. Finally

$$\log\Gamma(n) = \int_{-\infty}^0 \left[\frac{e^{\alpha n} - e^{\alpha}}{e^\alpha - 1} - (n-1)e^\alpha\right]\frac{d\alpha}{\alpha} \tag{19'}$$

if $1+\alpha$ be changed to $e^{-\alpha}$. The details of the reductions and the justification of the differentiation and integration will be left as exercises.

An approximate expression or, better, an *asymptotic expression*, that is, an expression with *small percentage error*, may be found for $\Gamma(n+1)$ when n is *large*. Choose the form (2) and note that the integrand $y^{2n+1}e^{-y^2}$ rises from 0 to a maximum at the point $y^2 = n + \frac{1}{2}$ and falls away again to 0. Make the change of variable $y = \sqrt{a} + w$, where $a = n + \frac{1}{2}$, so as to bring the origin under the maximum. Then

$$\Gamma(n+1) = 2\int_{-\sqrt{a}}^\infty (\sqrt{a} + w)^{2a}e^{-a-2\sqrt{a}w-w^2}dw,$$

or $$\Gamma(n+1) = 2a^a e^{-a}\int_{-\sqrt{a}}^\infty e^{2a\log\left(1+\frac{w}{\sqrt{a}}\right) - 2\sqrt{a}w - w^2}dw.$$

Now $$2a\log\left(1+\frac{w}{\sqrt{a}}\right) - 2\sqrt{a}w \leqq 0, \qquad -\sqrt{a} < w < \infty.$$

The integrand is therefore always less than e^{-w^2}, except when $w = 0$ and the integrand becomes 1. Moreover, as w increases, the integrand falls off very rapidly, and the chief part of the value of the integral may be obtained by integrating between rather narrow limits for w, say from -3 to $+3$. As α is large by hypothesis, the value of $\log\left(1 + w/\sqrt{\alpha}\right)$ may be obtained for small values of w from Maclaurin's Formula. Then

$$\Gamma(n+1) = 2\,\alpha^{\alpha}e^{-\alpha}\int_{-c}^{c} e^{-2w^2(1-\epsilon)}dw$$

is an approximate form for $\Gamma(n+1)$, where the quantity ϵ is about $\frac{2}{3}\,w/\sqrt{\alpha}$ and where the limits $\pm c$ of the integral are small relative to $\sqrt{\alpha}$. But as the integrand falls off so rapidly, there will be little error made in extending the limits to ∞ after dropping ϵ. Hence approximately

$$\Gamma(n+1) = 2\,\alpha^{\alpha}e^{-\alpha}\int_{-\infty}^{\infty} e^{-2w^2}dw = \sqrt{2\pi}\alpha^{\alpha}e^{-\alpha},$$

or
$$\Gamma(n+1) = \sqrt{2\pi}\,(n+\tfrac{1}{2})^{n+\frac{1}{2}}\,e^{-(n+\frac{1}{2})}(1+\eta), \tag{20}$$

where η is a small quantity approaching 0 as n becomes infinite.

EXERCISES

1. Establish the following formulas by changes of variable.

(α) $\Gamma(n) = \alpha^n\int_0^{\infty} x^{n-1}e^{-\alpha x}dx, \quad \alpha > 0,$ (β) $\int_0^{\frac{\pi}{2}}\sin^n x\,dx = \frac{1}{2}\,\mathrm{B}\left(\frac{n}{2}+\frac{1}{2}, \frac{1}{2}\right),$

(γ) $\mathrm{B}(n, n) = 2^{1-2n}\,\mathrm{B}(n, \frac{1}{2})$ by (6), (δ) $\int_0^1 x^{m-1}(1-x^2)^{n-1}dx = \frac{1}{2}\,\mathrm{B}(\frac{1}{2}m, n),$

(ϵ) $\displaystyle\int_0^1 \frac{x^{m-1}(1-x)^{n-1}}{(x+a)^{m+n}}dx = \frac{\mathrm{B}(m, n)}{a^n(1+a)^m} = \frac{1}{a^n(1+a)^m}\,\frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}$, take $\displaystyle\frac{x}{x+a} = \frac{y}{1+a}$,

(ζ) $\displaystyle\int_0^1 \frac{x^{m-1}(1-x)^{n-1}dx}{[ax+b(1-x)]^{m+n}} = \frac{\Gamma(m)\,\Gamma(n)}{a^m b^n\,\Gamma(m+n)}$, take $\displaystyle x = \frac{by}{a(1-y)+by}$,

(η) $\displaystyle\int_0^1 \frac{x^{m-1}(1-x)^{n-1}dx}{(b+cx)^{m+n}} = \frac{\mathrm{B}(m, n)}{b^n(b+c)^m},$ (θ) $\displaystyle\int_0^1 \frac{x^n dx}{\sqrt{1-x^2}} = \frac{\sqrt{\pi}}{2}\,\frac{\Gamma(\frac{1}{2}n+\frac{1}{2})}{\Gamma(\frac{1}{2}n+1)},$

(ι) $\displaystyle\int_0^1 x^m(1-x^n)^p dx = \frac{1}{n}\,\mathrm{B}\left(p+1, \frac{m+1}{n}\right),$ (κ) $\displaystyle\int_0^1 \frac{dx}{\sqrt{1-x^n}} = \frac{\sqrt{\pi}}{n}\,\frac{\Gamma(n^{-1})}{\Gamma(n^{-1}+\frac{1}{2})}.$

2. From $\Gamma(1) = 1$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ make a table of the values for every integer and half integer from 0 to 5 and plot the curve $y = \Gamma(x)$ from them.

3. By the aid of (10) and Ex. 1 (γ) prove the relations

$$\sqrt{\pi}\Gamma(2a) = 2^{2a-1}\Gamma(a)\,\Gamma(a+\tfrac{1}{2}), \qquad \sqrt{\pi}\Gamma(n) = 2^{n-1}\Gamma(\tfrac{1}{2}n)\,\Gamma(\tfrac{1}{2}n+\tfrac{1}{2}).$$

4. Given that $\Gamma(1.75) = 0.9191$, add to the table of Ex. 2 the values of $\Gamma(n)$ for every quarter from 0 to 3 and add the points to the plot.

5. With the aid of the Γ-function prove these relations (see Ex. 1):

$$(\alpha)\ \int_0^{\frac{\pi}{2}} \sin^n x\,dx = \int_0^{\frac{\pi}{2}} \cos^n x\,dx = \frac{1\cdot 3\cdot 5\cdots(n-1)}{2\cdot 4\cdot 6\cdots n}\,\frac{\pi}{2} \quad\text{or}\quad \frac{2\cdot 4\cdot 6\cdots(n-1)}{1\cdot 3\cdot 5\cdots n},$$

as n is even or odd.

$$(\beta)\ \int_0^1 \frac{x^{2n}dx}{\sqrt{1-x^2}} = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n}\,\frac{\pi}{2}, \qquad (\gamma)\ \int_0^1 \frac{x^{2n+1}dx}{\sqrt{1-x^2}} = \frac{2\cdot 4\cdot 6\cdots 2n}{1\cdot 3\cdot 5\cdots(2n+1)},$$

$$(\delta)\ \int_0^a x^2\sqrt{a^2-x^2}\,dx = \frac{\pi a^4}{16}, \qquad (\epsilon)\ \int_0^a x^2(a^2-x^2)^{\frac{3}{2}}dx = \frac{3\pi a^6}{96},$$

(ζ) Find $\displaystyle\int_0^1 \frac{dx}{\sqrt{1-x^4}}$ to four decimals. $\qquad(\eta)$ Find $\displaystyle\int_0^1 \frac{dx}{\sqrt{1-x^{\frac{1}{4}}}}$.

6. Find the areas of the quadrants of these curves:

(α) $x^{\frac{1}{2}} + y^{\frac{1}{2}} = a^{\frac{1}{2}}$, $\qquad(\beta)$ $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$, $\qquad(\gamma)$ $x^2 + y^{\frac{2}{3}} = 1$,

(δ) $x^2/a^2 + y^2/b^2 = 1$. $\qquad(\epsilon)$ the evolute $(ax)^{\frac{2}{3}} + (by)^{\frac{2}{3}} = (a^2-b^2)^{\frac{2}{3}}$.

7. Find centers of gravity and moments of inertia about the axes in Ex. 6.

8. Find volumes, centers of gravity, and moments of inertia of the octants of

(α) $x^{\frac{1}{2}} + y^{\frac{1}{2}} + z^{\frac{1}{2}} = a^{\frac{1}{2}}$, $\qquad(\beta)$ $x^{\frac{2}{3}} + y^{\frac{2}{3}} + z^{\frac{2}{3}} = a^{\frac{2}{3}}$, $\qquad(\gamma)$ $x^2 + y^2 + z^{\frac{2}{3}} = 1$.

9. (α) The sum of four proper fractions does not exceed unity; find the average value of their product. (β) The same if the sum of the squares does not exceed unity. (γ) What are the results in the case of k proper fractions?

10. Average $e^{-ax^2-by^2}$ under the supposition $ax^2 + by^2 \leqq H$.

11. Evaluate the definite integral (15′) by differentiation under the sign.

12. From (18) and $1 < \dfrac{\alpha}{1-e^{-\alpha}} < 1 + \alpha$ show that the magnitude of $D^2 \log\Gamma(n)$ is about $1/n$ for large values of n.

13. From Ex. 12, and Ex. 23, p. 76, show that the error in taking

$$\log\Gamma\left(n+\frac{1}{2}\right) \quad\text{for}\quad \int_n^{n+1}\log\Gamma(x)\,dx \quad\text{is about}\quad \frac{1}{24n+12}\log\Gamma\left(n+\frac{1}{2}\right).$$

14. Show that $\displaystyle\int_n^{n+1}\log\Gamma(x)\,dx = \int_0^1 \log\Gamma(n+x)\,dx$ and hence compare (15′), (20), and Ex. 13 to show that the small quantity η is about $(24n+12)^{-1}$.

15. Use a four-place table to find the logarithms of 5! and 10!. Find the logarithms of the approximate values by (20), and determine the percentage errors.

16. Assume $n = 11$ in (17) and evaluate the first integral. Take the logarithmic derivative of (20) to find an approximate expression for $\Gamma'(n)/\Gamma(n)$, and in particular compute the value for $n = 11$. Combine the results to find $\gamma = 0.578$. By more accurate methods it may be shown that Euler's Constant $\gamma = 0.577,215,665\cdots$.

17. Integrate (19′) from n to $n+1$ to find a definite integral for (15′). Subtract the integrals and add $\displaystyle\frac{1}{2}\log n = \int_{-\infty}^{0}\frac{e^{\alpha n}-e^{\alpha}}{2}\,\frac{d\alpha}{\alpha}$. Hence find

$$\log\Gamma(n) - n(\log n - 1) - \log\sqrt{2\pi} + \frac{1}{2}\log n = \int_{-\infty}^{0}\left[\frac{1}{e^{\alpha}-1} - \frac{1}{\alpha} + \frac{1}{2}\right]e^{\alpha n}\,\frac{d\alpha}{\alpha}.$$

18. Obtain *Stirling's approximation*, $\Gamma(n+1) = \sqrt{2\pi n}\, n^n e^{-n}$, either by comparing it with the one already found or by applying the method of the text, with the substitution $x = n + \sqrt{2n}\, y$, to the original form (1) of $\Gamma(n+1)$.

19. The relation $\displaystyle\prod_{k=1}^{k=n-1} \sin\frac{k\pi}{n} = \sin\frac{\pi}{n}\sin\frac{2\pi}{n}\cdots\sin\frac{(n-1)\pi}{n} = \frac{n}{2^{n-1}}$ may be obtained from the roots of unity (§ 72); for $x^n - 1 = (x-1)\prod\left(x - e^{-\frac{2k\pi i}{n}}\right)$,

$$n = \lim_{x \doteq 1}\frac{x^n - 1}{x - 1} = \prod_{k=1}^{k=n-1}\left(1 - e^{-\frac{2k\pi i}{n}}\right), \qquad \prod_{k=1}^{k=n-1}\frac{e^{\frac{k\pi i}{n}}}{2i} = \frac{e^{(n-1)\frac{\pi i}{2}}}{(2i)^{n-1}} = \frac{1}{2^{n-1}}.$$

150. The error function. Suppose that measurements to determine the magnitude of a certain object be made, and let $m_1, m_2, \cdots, m_n$ be a set of n determinations each made independently of the other and each worthy of the same weight. Then the quantities

$$q_1 = m_1 - m, \qquad q_2 = m_2 - m, \qquad \cdots, \qquad q_n = m_n - m,$$

which are the differences between the observed values and the assumed value m, are the errors committed; their sum is

$$q_1 + q_2 + \cdots + q_n = (m_1 + m_2 + \cdots + m_n) - mn.$$

It will be taken as a fundamental axiom that on the average the errors in excess, the positive errors, and the errors in defect, the negative errors, are evenly balanced so that their sum is zero. In other words it will be assumed that the mean value

$$nm = m_1 + m_2 + \cdots + m_n \quad \text{or} \quad m = \frac{1}{n}(m_1 + m_2 + \cdots + m_n) \tag{21}$$

is the most probable value for m as determined from $m_1, m_2, \cdots, m_n$. Note that the average value m is that which makes the sum of the squares of the errors a minimum; hence the term "least squares."

Before any observations have been taken, the chance that any particular error q should be made is 0, and the chance that an error lie within infinitesimal limits, say between q and $q + dq$, is infinitesimal; let the chance be assumed to be a function of the size of the error, and write $\phi(q)dq$ as the chance that an error lie between q and $q + dq$. It may be seen that $\phi(q)$ may be expected to decrease as q increases; for, under the reasonable hypothesis that an observer is not so likely to be far wrong as to be somewhere near right, the chance of making an error between 8.0 and 8.1 would be less than that of making an error between 1.0 and 1.1. The function $\phi(q)$ is called the error function. It will be said that the chance of making an error q_i is $\phi(q_i)$; to put it more precisely, this means simply that $\phi(q_i)dq$ is the chance of making an error which lies between q_i and $q_i + dq$.

It is a fundamental principle of the theory of chance that the chance that several independent events take place is the product of the chances for each separate event. The probability, then, that the errors $q_1, q_2, \cdots, q_n$ be made is the product

$$\phi(q_1)\,\phi(q_2)\cdots\phi(q_n) = \phi(m_1 - m)\,\phi(m_2 - m)\cdots\phi(m_n - m). \quad (22)$$

The fundamental axiom (21) is that this probability is a maximum when m is the arithmetic mean of the measurements $m_1, m_2, \cdots, m_n$; for the errors, measured from the mean value, are on the whole less than if measured from some other value.* If the probability is a maximum, so is its logarithm; and the derivative of the logarithm of (22) with respect to m is

$$\frac{\phi'(m_1 - m)}{\phi(m_1 - m)} + \frac{\phi'(m_2 - m)}{\phi(m_2 - m)} + \cdots + \frac{\phi'(m_n - m)}{\phi(m_n - m)} = 0$$

when $\quad + q_2 + \cdots + q_n = (m_1 - m) + (m_2 - m) + \cdots + (m_n - m) = 0.$

It remains to determine ϕ from these relations.

For brevity let $F(q)$ be the function $F = \phi'/\phi$ which is the ratio of $\phi'(q)$ to $\phi(q)$. Then the conditions become

$$F(q_1) + F(q_2) + \cdots + F(q_n) = 0 \quad \text{when} \quad q_1 + q_2 + \cdots + q_n = 0.$$

In particular if there are only two observations, then

$$F(q_1) + F(q_2) = 0 \quad \text{and} \quad q_1 + q_2 = 0 \quad \text{or} \quad q_2 = -q_1.$$

Then $\quad F(q_1) + F(-q_1) = 0 \quad \text{or} \quad F(-q) = -F(q).$

Next if there are three observations, the results are

$$F(q_1) + F(q_2) + F(q_3) = 0 \quad \text{and} \quad q_1 + q_2 + q_3 = 0.$$

Hence $\quad F(q_1) + F(q_2) = -F(q_3) = F(-q_3) = F(q_1 + q_2).$

Now from $\quad F(x) + F(y) = F(x + y)$

the function F may be determined (Ex. 9, p. 45) as $F(x) = Cx$. Then

$$F(q) = \frac{\phi'(q)}{\phi(q)} = Cq, \qquad \log\phi(q) = \frac{1}{2}Cq^2 + K,$$

and $\quad \phi(q) = e^{\frac{1}{2}Cq^2 + K} = Ge^{\frac{1}{2}Cq^2}.$

This determination of ϕ contains two arbitrary constants which may be further determined. In the first place, note that C is negative, for $\phi(q)$ decreases as q increases. Let $\frac{1}{2}C = -k^2$. In the second place, the

* The derivation of the expression for ϕ is physical rather than logical in its argument. The real justification or proof of the validity of the expression obtained is *à posteriori* and depends on the experience that in practice errors do follow the law (24).

error q must lie within the interval $-\infty < q < +\infty$ which comprises all possible values. Hence

$$\int_{-\infty}^{+\infty}\phi(q)\,dq = 1, \qquad G\int_{-\infty}^{+\infty}e^{-k^2q^2}dq = 1. \tag{23}$$

For the chance that an error lie between q and $q + dq$ is ϕdq, and if an interval $a \leqq q \leqq b$ be given, the chance of an error in it is

$$\sum_a^b \phi(q)\,dq \quad \text{or, better,} \quad \lim \sum_a^b \phi(q)\,dq = \int_a^b \phi(q)\,dq,$$

and finally the chance that $-\infty < q < +\infty$ represents a certainty and is denoted by 1. The integral (23) may be evaluated (§ 143). Then $G\sqrt{\pi}/k = 1$ and $G = k/\sqrt{\pi}$. Hence *

$$\phi(q) = \frac{k}{\sqrt{\pi}}e^{-k^2q^2}. \tag{24}$$

The remaining constant k is essential; it measures the accuracy of the observer. If k is large, the function $\phi(q)$ falls very rapidly from the large value $k/\sqrt{\pi}$ for $q = 0$ to very small values, and it appears that the observer is far more likely to make a small error than a large one; but if k is small, the function ϕ falls very slowly from its value $k/\sqrt{\pi}$ for $q = 0$ and denotes that the observer is almost as likely to make reasonably large errors as small ones.

151. If only the numerical value be considered, the probability that the error lie numerically between q and $q + dq$ is

$$\frac{2k}{\sqrt{\pi}}e^{-k^2q^2}dq, \quad \text{and} \quad \frac{2k}{\sqrt{\pi}}\int_0^\xi e^{-k^2q^2}dq$$

is the chance that an error be numerically less than ξ. Now

$$\psi(\xi) = \frac{2k}{\sqrt{\pi}}\int_0^\xi e^{-k^2q^2}dq = \frac{2}{\sqrt{\pi}}\int_0^{k\xi}e^{-x^2}dx \tag{25}$$

is a function defined by an integral with a variable upper limit, and the problem of computing the value of the function for any given value of ξ reduces to the problem of computing the integral. The integrand may be expanded by Maclaurin's Formula

$$e^{-x^2} = 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \frac{x^8}{4!} - \frac{x^{10}e^{-\theta x^2}}{5!}, \qquad 0 < \theta < 1,$$

$$\int_0^x e^{-x^2}dx = x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - R, \qquad R < \frac{x^{11}}{1320}. \tag{26}$$

* The reader may now verify the fact that, with ϕ as in (24), the product (22) is a maximum if the sum of the squares of the errors is a minimum as demanded by (21).

For small values of x this series is satisfactory; for $x \leqq \frac{1}{2}$ it will be accurate to five decimals.

The *probable error* is the technical term used to denote that error ξ which makes $\psi(\xi) = \frac{1}{2}$; that is, the error such that the chance of a smaller error is $\frac{1}{2}$ and the chance of a larger error is also $\frac{1}{2}$. This is found by solving for x the equation

$$\frac{\sqrt{\pi}}{2} \cdot \frac{1}{2} = 0.44311 = \int_0^x e^{-x^2} dx = x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216}.$$

The first term alone indicates that the root is near $x = .45$, and a trial with the first three terms in the series indicates the root as between $x = .47$ and $x = .48$. With such a close approximation it is easy to fix the root to four places as

$$x = k\xi = 0.4769 \quad \text{or} \quad \xi = 0.4769\, k^{-1}. \tag{27}$$

That the probable error should depend on k is obvious.

For large values of $x = k\xi$ the method of expansion by Maclaurin's Formula is a very poor one for calculating $\psi(\xi)$; too many terms are required. It is therefore important to obtain an *expansion according to descending powers of x*. Now

$$\int_0^x e^{-x^2} dx = \int_0^\infty e^{-x^2} dx - \int_x^\infty e^{-x^2} dx = \frac{1}{2}\sqrt{\pi} - \int_x^\infty e^{-x^2} dx$$

and
$$\int_x^\infty e^{-x^2} dx = \int_x^\infty \frac{1}{x} x e^{-x^2} dx = \left[-\frac{e^{-x^2}}{2x}\right]_x^\infty - \frac{1}{2}\int_x^\infty \frac{e^{-x^2} dx}{x^2}.$$

The limits may be substituted in the first term and the method of integration by parts may be applied again. Thus

$$\begin{aligned}\int_x^\infty e^{-x^2} dx &= \frac{e^{-x^2}}{2x}\left(1 - \frac{1}{2x^2}\right) + \frac{1 \cdot 3}{2^2}\int_x^\infty \frac{e^{-x^2} dx}{x^4} \\ &= \frac{e^{-x^2}}{2x}\left(1 - \frac{1}{2x^2} + \frac{1 \cdot 3}{2^2 x^4}\right) - \frac{1 \cdot 3 \cdot 5}{2^3}\int_x^\infty \frac{e^{-x^2} dx}{x^6},\end{aligned}$$

and so on indefinitely. It should be noticed, however, that the term

$$T = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n x^{2n}} \frac{e^{-x^2}}{2x} \text{ diverges as } n = \infty.$$

In fact although the denominator is multiplied by $2x^2$ at each step, the numerator is multiplied by $2n - 1$, and hence after the integrations by parts have been applied so many times that $n > x^2$ the terms in the parenthesis begin to increase. It is worse than useless to carry the integrations further. The integral which remains is (Ex. 5, p. 29)

$$\frac{1\cdot 3\cdot 5\cdots(2n+1)}{2^{n+1}}\int_x^\infty \frac{e^{-x^2}dx}{x^{2n+2}} < \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^{n-1}x^{2n-1}}e^{-x^2} < T.$$

Thus the integral is less than the last term of the parenthesis, and it is possible to write the *asymptotic series*

$$\int_0^x e^{-x^2}dx = \frac{1}{2}\sqrt{\pi} - \frac{e^{-x^2}}{2x}\left(1 - \frac{1}{2x^2} + \frac{1\cdot 3}{2^2x^4} - \frac{1\cdot 3\cdot 5}{2^3x^6} + \cdots\right) \qquad (28)$$

with the assurance that *the value obtained by using the series will differ from the true value by less than the last term which is used in the series.* This kind of series is of frequent occurrence.

In addition to the probable error, the *average numerical error* and the *mean square error*, that is, the average of the square of the error, are important. In finding the averages the probability $\phi(q)\,dq$ may be taken as the weight; in fact the probability is in a certain sense the simplest weight because the sum of the weights, that is, the sum of the probabilities, is 1 if an average over the whole range of possible values is desired. For the average numerical error and mean square error

$$\begin{aligned} \overline{|q|} &= \frac{2k}{\sqrt{\pi}}\int_0^\infty qe^{-k^2q^2}dq = \frac{1}{k\sqrt{\pi}} = \frac{0.5643}{k}, \\ \overline{q^2} &= \frac{2k}{\sqrt{\pi}}\int_0^\infty q^2e^{-k^2q^2}dq = \frac{1}{2k^2}, \qquad \sqrt{\overline{q^2}} = \frac{0.7071}{k}. \end{aligned} \qquad (29)$$

It is seen that the average error is greater than the probable error, and that the square root of the mean square error is still larger. In the case of a given set of n observations the averages may actually be computed as

$$\overline{|q|} = \frac{|q_1| + |q_2| + \cdots + |q_n|}{n} = \frac{1}{k\sqrt{\pi}}, \qquad k = \frac{1}{\overline{|q|}\sqrt{\pi}},$$

$$\overline{q^2} = \frac{q_1^2 + q_2^2 + \cdots + q_n^2}{n} = \frac{1}{2k^2}, \qquad k = \frac{1}{\sqrt{\overline{q^2}}\sqrt{2}}.$$

Moreover, $\qquad \pi\,\overline{|q|}^2 = 2\,\overline{q^2}.$

It cannot be expected that the two values of k thus found will be precisely equal or that the last relation will be exactly fulfilled; but so well does the theory of errors represent what actually arises in practice that unless the two values of k are nearly equal and the relation nearly satisfied there are fair reasons for suspecting that the observations are not bona fide.

152. Consider the question of the application of these theories to the errors made in rifle practice on a target. Here there are two

errors, one due to the fact that the shots may fall to the right or left of the central vertical, the other to their falling above or below the central horizontal. In other words, each of the coördinates (x, y) of the position of a shot will be regarded as subject to the law of errors independently of the other. Then

$$\frac{k}{\sqrt{\pi}} e^{-k^2x^2}dx, \qquad \frac{k'}{\sqrt{\pi}} e^{-k'^2y^2}dy, \qquad \frac{kk'}{\pi} e^{-k^2x^2-k'^2y^2}dxdy$$

will be the probabilities that a shot fall in the vertical strip between x and $x + dx$, in the horizontal strip between y and $y + dy$, or in the small rectangle common to the two strips. Moreover it will be assumed that the accuracy is the same with respect to horizontal and vertical deviations, so that $k = k'$.

These assumptions may appear too special to be reasonable. In particular it might seem as though the accuracies in the two directions would be very different, owing to the possibility that the marksman's aim should tremble more to the right and left than up and down, or vice versa, so that $k \neq k'$. In this case the shots would not tend to lie at equal distances in all directions from the center of the target, but would dispose themselves in an elliptical fashion. Moreover as the shooting is done from the right shoulder it might seem as though there would be some inclined line through the center of the target along which the accuracy would be least, and a line perpendicular to it along which the accuracy would be greatest, so that the disposition of the shots would not only be elliptical but inclined. To cover this general assumption the probability would be taken as

$$Ge^{-k^2x^2-2\lambda xy-k'^2y^2}dxdy, \quad \text{with} \quad G\int_{-\infty}^{\infty}\int e^{-k^2x^2-2\lambda xy-k'^2y^2}dxdy = 1$$

as the condition that the shots lie somewhere. See the exercises below.

With the special assumptions, it is best to transform to polar coördinates. The important quantities to determine are the average distance of the shots from the center, the mean square distance, the probable distance, and the most probable distance. It is necessary to distinguish carefully between the probable distance, which is by definition the distance such that half the shots fall nearer the center and half fall farther away, and the most probable distance, which by definition is that distance which occurs most frequently, that is, the distance of the ring between r and $r + dr$ in which most shots fall.

The probability that the shot lies in the element $rdrd\phi$ is

$$\frac{k^2}{\pi} e^{-k^2r^2}rdrd\phi, \quad \text{and} \quad 2k^2e^{-k^2r^2}rdr,$$

obtained by integrating with respect to ϕ, is the probability that the shot lies in the ring from r to $r + dr$. The *most probable* distance r_p is

that which makes this a maximum, that is,

$$\frac{d}{dr}(e^{-k^2r^2}r) = 0 \quad \text{or} \quad r_p = \frac{1}{\sqrt{2}\,k} = \frac{0.7071}{k}. \tag{30}$$

The *mean* distance and the *mean square* distance are respectively

$$\begin{aligned} \bar{r} &= \int_0^\infty 2\,k^2 e^{-k^2r^2} r^2 dr = \frac{\sqrt{\pi}}{2\,k}, & \bar{r} &= \frac{0.8862}{k}, \\ \overline{r^2} &= \int_0^\infty 2\,k^2 e^{-k^2r^2} r^3 dr = \frac{1}{k^2}, & \sqrt{\overline{r^2}} &= \frac{1.0000}{k}. \end{aligned} \tag{30'}$$

The *probable* distance r_ξ is found by solving the equation

$$\frac{1}{2} = \int_0^{r_\xi} 2\,k^2 e^{-k^2r^2} r\,dr = 1 - e^{-k^2r_\xi^2}, \qquad r_\xi = \frac{\sqrt{\log 2}}{k} = \frac{0.8326}{k}. \tag{30''}$$

Hence $$r_p < r_\xi < \bar{r} < \sqrt{\overline{r^2}}.$$

The chief importance of these considerations lies in the fact that, owing to Maxwell's assumption, analogous considerations may be applied to the velocities of the molecules of a gas. Let u, v, w be the component velocities of a molecule in three perpendicular directions so that $V = (u^2 + v^2 + w^2)^{\frac{1}{2}}$ is the actual velocity. The assumption is made that the individual components u, v, w obey the law of errors. The probability that the components lie between the respective limits u and $u + du$, v and $v + dv$, w and $w + dw$ is

$$\frac{k^3}{\pi\sqrt{\pi}} e^{-k^2u^2 - k^2v^2 - k^2w^2} du\,dv\,dw, \quad \text{and} \quad \frac{k^3}{\pi\sqrt{\pi}} e^{-k^2V^2} V^2 \sin\theta\, dV\, d\theta\, d\phi$$

is the corresponding expression in polar coördinates. There will then be a most probable, a probable, a mean, and a mean square velocity. Of these, the last corresponds to the mean kinetic energy and is subject to measurement.

EXERCISES

1. If $k = 0.04475$, find to three places the probability of an error $\xi < 12$.

2. Compute $\int_0^x e^{-x^2}dx$ to three places for (α) $x = 0.2$, (β) $x = 0.8$.

3. State how many terms of (28) should be taken to obtain the best value for the integral to $x = 2$ and obtain that value.

4. How accurately will (28) determine $\int_0^4 e^{-x^2}dx - \frac{1}{2}\sqrt{\pi}$? Compute.

5. Obtain these asymptotic expansions and extend them to find the general law. Show that the error introduced by omitting the integral is less than the last term retained in the series. Show further that the general term diverges when n becomes infinite.

$$(\alpha) \int_0^x \cos x^2 dx = \frac{1}{2}\sqrt{\frac{\pi}{2}} + \frac{\sin x^2}{2x} - \frac{\cos x^2}{2^2 x^3} + \frac{1\cdot 3}{2^2}\int_x^\infty \cos x^2 \frac{dx}{x^4},$$

$$(\beta) \int_0^x \sin x^2 dx = \frac{1}{2}\sqrt{\frac{\pi}{2}} - \frac{\cos x^2}{2x} - \frac{\sin x^2}{2^2 x^3} + \frac{1\cdot 3}{2^2}\int_x^\infty \sin x^2 \frac{dx}{x^4},$$

$$(\gamma) \int_0^x \frac{\sin x}{x} dx, \ x \text{ large}. \qquad (\delta) \int_0^x \left(\frac{\sin x}{x}\right)^2 dx, \ x \text{ large}.$$

6. (α) Find the value of the average of any odd power $2n+1$ of the error; (β) also for the average of any even power; (γ) also for any power.

7. The observations 195, 225*, 190, 210, 205, 180*, 170*, 190, 200, 210, 210, 220*, 175*, 192 were obtained for deflections of a galvanometer. Compute k from the mean error and mean square error and compare the results. Suppose the observations marked *, which show great deviations, were discarded; compute k by the two methods and note whether the agreement is so good.

8. Find the average value of the product qq' of two errors selected at random and the average of the product $|q|\cdot|q'|$ of numerical values.

9. Show that the various velocities for a gas are $V_p = \frac{1}{k}$, $V_\xi = \frac{1.0875}{k}$, $\bar{V} = \frac{2}{\sqrt{\pi}k} = \frac{1.1284}{k}$, $\sqrt{\overline{V^2}} = \frac{\sqrt{3}}{\sqrt{2}k} = \frac{1.2247}{k}$.

10. For oxygen (at 0° C. and 76 cm. Hg.) the square root of the mean square velocity is 462.2 meters per second. Find k and show that only about 13 or 14 molecules to the thousand are moving as slow as 100 m./sec. What speed is most probable?

11. Under the general assumption of ellipticity and inclination in the distribution of the shots show that the area of the ellipse $k^2x^2 + 2\lambda xy + k'^2y^2 = H$ is $\pi H(k^2k'^2 - \lambda^2)^{-\frac{1}{2}}$, and the probability may be written $Ge^{-H}\pi(k^2k'^2-\lambda^2)^{-\frac{1}{2}}dH$.

12. From Ex. 11 establish the relations $(\alpha)\ G = \frac{1}{\pi}\sqrt{k^2k'^2 - \lambda^2}$,

$$(\beta)\ \overline{x^2} = \frac{k'^2}{2(k^2k'^2-\lambda^2)}, \qquad (\gamma)\ \overline{y^2} = \frac{k^2}{2(k^2k'^2-\lambda^2)}, \qquad (\delta)\ \overline{xy} = \frac{-\lambda}{2(k^2k'^2-\lambda^2)}.$$

13. Find H_p, $H_\xi = 0.693$, $\bar{H}$, $\overline{H^2}$ in the above problem.

14. Take 20 measurements of some object. Determine k by the two methods and compare the results. Test other points of the theory.

153. Bessel functions. The use of a definite integral to define functions which satisfy a given differential equation may be illustrated by the treatment of $xy'' + (2n+1)y' + xy = 0$, which at the same time will afford a new investigation of some functions which have previously been briefly discussed (§§ 107–108). To obtain a solution of this equation, or of any equation, in the form of a definite integral, some special type of integrand is assumed in part and the remainder of the

integrand and the limits for the integral are then determined so that the equation is satisfied. In this case try the form

$$y(x) = \int e^{ixt}Tdt, \qquad y' = \int ite^{ixt}Tdt, \qquad y'' = \int -t^2e^{ixt}Tdt,$$

where T is a function of t, and the derivatives are found by differentiating under the sign. Integrate y and y'' by parts and substitute in the equation. Then

$$(1-t^2)\,Te^{ixt}\big] - \int e^{ixt}[T'(1-t^2) + (2n-1)tT]dt = 0,$$

where the bracket after the first term means that the difference of the values for the upper and lower limit of the integral are to be taken; these limits and the form of T remain to be determined so that the expression shall really be zero.

The integral may be made to vanish by so choosing T that the bracket vanishes; this calls for the integration of a simple differential equation. The result then is

$$T = (1-t^2)^{n-\frac{1}{2}}, \qquad (1-t^2)^{n+\frac{1}{2}}e^{ixt}\big] = 0.$$

The integral vanishes, and the integrated term will vanish provided $t = \pm 1$ or $e^{ixt} = 0$. If x be assumed to be real and positive, the exponential will approach 0 when $t = 1 + iK$ and K becomes infinite. Hence

$$y(x) = \int_{-1}^{+1} e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt \quad \text{and} \quad z(x) = \int_{+1}^{1+i\infty} e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt \quad (31)$$

are solutions of the differential equation. In the first the integral is an infinite integral when $n < +\frac{1}{2}$ and fails to converge when $n \leqq -\frac{1}{2}$. The solution is therefore defined only when $n > -\frac{1}{2}$. The second integral is always an infinite integral because one limit is infinite. The examination of the integrals for uniformity is found below.

Consider $\int_{-1}^{+1} e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt$ with $n < \frac{1}{2}$ so that the integral is infinite.

$$\int_{-1}^{+1} e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt = \int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}\cos xtdt + i\int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}\sin xtdt.$$

From considerations of symmetry the second integral vanishes. Then

$$\left|\int_{-1}^{+1} e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt\right| = \left|\int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}\cos xtdt\right| \leqq \int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}dt.$$

This last integral with a positive integrand converges when $n > -\frac{1}{2}$, and hence the given integral converges uniformly for all values of x and defines a continuous function. The successive differentiations under the sign give the results

$$-\int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}t\sin xtdt, \qquad -\int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}t^2\cos xtdt.$$

These integrals also converge uniformly, and hence the differentiations were justifiable. The second integral (31) may be written with $t = 1 + iu$, as

$$\left| i\int_{u=0}^{\infty} e^{ix(1+iu)}(1-\overline{1+iu}^2)^{n-\frac{1}{2}}du \right| \leqq \int_0^{\infty} e^{-ux}(4u^2+u^4)^{\frac{1}{2}n-\frac{1}{4}}du.$$

This integral converges for all values of $x > 0$ and $n > -\frac{1}{2}$. Hence the given integral converges uniformly for all values of $x \geqq x_0 > 0$, and defines a continuous function; when $x = 0$ it is readily seen that the integral diverges and could not define a continuous function. It is easy to justify the differentiations as before.

The first form of the solution may be expanded in series.

$$y(x) = \int_{-1}^{+1} e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt = \int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}\cos xtdt$$

$$= 2\int_0^1 (1-t^2)^{n-\frac{1}{2}}\cos xtdt \tag{32}$$

$$= 2\int_0^1 (1-t^2)^{n-\frac{1}{2}}\left(1-\frac{x^2t^2}{2!}+\frac{x^4t^4}{4!}-\frac{x^6t^6}{6!}+\theta\frac{x^8t^8}{8!}\right)dt, \quad 0 < |\theta| < 1.$$

The expansion may be carried to as many terms as desired. Each of the terms separately may be integrated by B- or Γ-functions.

$$2\int_0^1 (1-t^2)^{n-\frac{1}{2}}\frac{x^{2k}t^{2k}}{2k!} = 2\frac{x^{2k}}{\Gamma(2k+1)}\int_0^{\frac{\pi}{2}}\sin^{2n}\phi\cos^{2k}\phi d\phi$$

$$= \frac{x^{2k}\Gamma(n+\frac{1}{2})\Gamma(k+\frac{1}{2})}{\Gamma(2k+1)\Gamma(n+k+1)} = \frac{x^{2k}\Gamma(n+\frac{1}{2})\sqrt{\pi}}{2^{2k}\Gamma(k+1)\Gamma(n+k+1)},$$

and $$J_n(x) = \frac{x^n y(x)}{2^n\sqrt{\pi}\,\Gamma(n+\frac{1}{2})} = \sum_{k=0}\frac{(-1)^k x^{n+2k}}{2^{n+2k}\Gamma(k+1)\Gamma(n+k+1)} \tag{33}$$

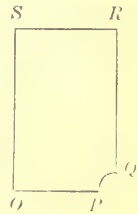
is then taken as the definition of the special function $J_n(x)$, where the expansion may be carried as far as desired, with the coefficient θ for the last term. If n is an integer, the Γ-functions may be written as factorials.

154. The second solution of the differential equation, namely

$$z(x) = y_1(x) + iy_2(x) = \int_1^{1+i\infty} -2e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt, \tag{31'}$$

where the coefficient -2 has been inserted for convenience, is for some purposes more useful than the first. It is complex, and, as the equation is real and x is taken as real, it affords two solutions, namely its real part and its pure imaginary part, each of which must satisfy the equation. As $y(x)$ converges for $x = 0$ and $z(x)$ diverges for $x = 0$, so that $y_1(x)$ or

$y_2(x)$ diverges, it follows that $y(x)$ and $y_1(x)$ or $y(x)$ and $y_2(x)$ must be independent; and as the equation can have but two independent solutions, one of the pairs of solutions must constitute a complete solution. It will now be shown that $y_1(x) = y(x)$ and that $Ay(x) + By_2(x)$ is therefore the complete solution of $xy'' + (2n+1)y' + xy = 0$.



Consider the line integral around the contour $0, 1-\epsilon, 1+\epsilon i, 1+\infty i, \infty i, 0$, or $OPQRS$. As the integrand has a continuous derivative at every point on or within the contour, the integral is zero (§ 124). The integrals along the little quadrant PQ and the unit line RS at infinity may be made as small as desired by taking the quadrant small enough and the line far enough away. The integral along SO is pure imaginary, namely, with $t = iu$,

$$\int_{SO} -2e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt = 2i\int_0^S e^{-xu}(1+u^2)^{n-\frac{1}{2}}du.$$

The integral along OP is complex, namely

$$\int_{OP} -2e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt$$
$$= -2\int_0^P (1-t^2)^{n-\frac{1}{2}}\cos xt\,dt - 2i\int_0^P (1-t^2)^{n-\frac{1}{2}}\sin xt\,dt.$$

Hence $$0 = -2\int_0^P (1-t^2)^{n-\frac{1}{2}}\cos xt\,dt - 2i\int_0^P (1-t^2)^{n-\frac{1}{2}}\sin xt\,dt + \zeta_1$$
$$+\int_Q^R -2e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt + \zeta_2 + 2i\int_0^S e^{-xu}(1+u^2)^{n-\frac{1}{2}}du,$$

where ζ_1 and ζ_2 are small. Equate real and imaginary parts to zero separately after taking the limit.

$$2\int_0^1 (1-t^2)^{n-\frac{1}{2}}\cos xt\,dt = y(x) = \mathcal{R}\int_1^{1+i\infty} -2e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt = y_1(x).$$

$$2\int_0^1 (1-t^2)^{n-\frac{1}{2}}\sin xt\,dt - 2\int_0^\infty e^{-xu}(1+u^2)^{n-\frac{1}{2}}du$$
$$= \mathcal{I}\int_1^{1+i\infty} -2e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt = y_2(x).$$

The signs $\mathcal{R}$ and $\mathcal{I}$ are used to denote respectively real and imaginary parts. The identity of $y(x)$ and $y_1(x)$ is established and the new solution $y_2(x)$ is found as a difference of two integrals.

It is now possible to obtain the important expansion of the solutions $y(x)$ and $y_2(x)$ in *descending* powers of x. For

$$\int_1^{1+i\infty} -2e^{ixt}(1-t^2)^{n-\frac{1}{2}}dt = \int_0^\infty -2ie^{ix-ux}(u^2-2iu)^{n-\frac{1}{2}}du, \quad t=1+iu.$$

Since $x \neq 0$, the transformation $ux = v$ is permissible and gives

$$2^{n+\frac{1}{2}}(-i)^{n+\frac{1}{2}}e^{ix}x^{-n-\frac{1}{2}}\int_0^\infty e^{-v}v^{n-\frac{1}{2}}\left(1+\frac{vi}{2x}\right)^{n-\frac{1}{2}}dv$$

$$= 2^{n+\frac{1}{2}}x^{-n-\frac{1}{2}}e^{i\left[x-\left(n+\frac{1}{2}\right)\frac{\pi}{2}\right]}\int_0^\infty e^{-v}v^{n-\frac{1}{2}}\times$$

$$\left(1+\frac{n-\frac{1}{2}}{2x}vi-\frac{(n-\frac{1}{2})(n-\frac{3}{2})}{2!\,(2x)^2}v^2-\cdots\right)dv.$$

The expansion by the binomial theorem may be carried as far as desired; but as the integration is subsequently to be performed, the values of v must be allowed a range from 0 to ∞ and the use of Taylor's Formula with a remainder is required — the series would not converge. The result of the integration is

$$z(x) = 2^{n+\frac{1}{2}}x^{-n-\frac{1}{2}}\Gamma(n+\tfrac{1}{2})e^{i\left[x-\left(n+\frac{1}{2}\right)\frac{\pi}{2}\right]}[P(x)+iQ(x)], \qquad (34)$$

where $$Q(x) = \frac{n^2-\frac{1}{4}}{2x} - \frac{(n^2-\frac{1}{4})(n^2-\frac{9}{4})(n^2-\frac{25}{4})}{3!\,(2x)^3} + \cdots,$$

$$P(x) = 1 - \frac{(n^2-\frac{1}{4})(n^2-\frac{9}{4})}{2!\,(2x)^2} + \frac{(n^2-\frac{1}{4})(n^2-\frac{9}{4})(n^2-\frac{25}{4})(n^2-\frac{49}{4})}{4!\,(2x)^4} - \cdots.$$

Take real and imaginary parts and divide by $2^n x^{-n}\sqrt{\pi}\,\Gamma(n+\frac{1}{2})$. Then

$$J_n(x) = \sqrt{\frac{2}{\pi x}}\left[P(x)\cos\left(x-\left(n+\frac{1}{2}\right)\frac{\pi}{2}\right) - Q(x)\sin\left(x-\left(n+\frac{1}{2}\right)\frac{\pi}{2}\right)\right],$$

$$K_n(x) = \sqrt{\frac{2}{\pi x}}\left[Q(x)\cos\left(x-\left(n+\frac{1}{2}\right)\frac{\pi}{2}\right) + P(x)\sin\left(x-\left(n+\frac{1}{2}\right)\frac{\pi}{2}\right)\right]$$

are two independent Bessel functions which satisfy the equation (35) of § 107. If $n+\frac{1}{2}$ is an integer, P and Q terminate and the solutions are expressed in terms of elementary functions (§ 108); but if $n+\frac{1}{2}$ is not an integer, P and Q are merely asymptotic expressions which do not terminate of themselves, but must be cut short with a remainder term because of their tendency to diverge after a certain point; for tolerably large values of x and small values of n the values of $J_n(x)$ and $K_n(x)$ may, however, be computed with great accuracy by using the first few terms of P and Q.

The integration to find P and Q offers no particular difficulty.

$$\int_0^\infty e^{-v}v^{n-\frac{1}{2}+k}dv = \Gamma(n+\tfrac{1}{2}+k) = (n+k-\tfrac{1}{2})(n+k-\tfrac{3}{2})\cdots(n+\tfrac{1}{2})\Gamma(n+\tfrac{1}{2}).$$

The factors previous to $\Gamma(n+\frac{1}{2})$ combine with $n-\frac{1}{2}, n-\frac{3}{2}, \cdots, n-k+\frac{1}{2}$, which occur in the kth term of the binomial expansion and give the numerators of the terms in P and Q. The remainder term must, however, be discussed. The integral form (p. 57) will be used.

$$R_k = \int_0^v \frac{t^{k-1}}{(k-1)!} f^{(k)}(v-t)\,dt,$$

$$f^{(k)} = \left(n-\frac{1}{2}\right)\cdots\left(n-k+\frac{1}{2}\right)\left(\frac{i}{2x}\right)^k\left(1+\frac{vi}{2x}\right)^{n-k-\frac{1}{2}}.$$

Let it be supposed that the expansion has been carried so far that $n-k-\frac{1}{2}<0$. Then $(1+vi/2x)^{n-k-\frac{1}{2}}$ is numerically greatest when $v=0$ and is then equal to 1. Hence

$$|R_k| < \int_0^v \frac{t^{k-1}}{(k-1)!}\frac{|(n-\frac{1}{2})\cdots(n-k+\frac{1}{2})|}{(2x)^k}dt = \frac{v^k}{k!}\frac{|(n-\frac{1}{2})\cdots(n-k+\frac{1}{2})|}{(2x)^k},$$

and $$\left|\int_0^\infty e^{-v}v^{n-\frac{1}{2}}R_k dv\right| < \frac{\left|\left(n^2-\frac{1}{4}\right)\cdots\left(n^2-\frac{(2k-1)^2}{4}\right)\right|}{k!\,(2x)^k}\Gamma\left(n+\frac{1}{2}\right).$$

It therefore appears that when $k>n-\frac{1}{2}$ the error made in neglecting the remainder is less than the last term kept, and for the maximum accuracy the series for $P+iQ$ should be broken off between the least term and the term just following.

EXERCISES

1. Solve $xy''+(2n+1)y'-xy=0$ by trying Te^{xt} as integrand.

$$A\int_{-1}^{+1}(1-t^2)^{n-\frac{1}{2}}e^{xt}dt + B\int_{-\infty}^{-1}(t^2-1)^{n-\frac{1}{2}}e^{xt}dt, \qquad x>0, \qquad n>-\tfrac{1}{2}.$$

2. Expand the first solution in Ex. 1 into series; compare with $y(ix)$ above.

3. Try $T(1-tx)^m$ on $x(1-x)y''+[\gamma-(\alpha+\beta+1)x]y'-\alpha\beta y=0$.

One solution is $\displaystyle\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tx)^{-\alpha}dt$. $\qquad \beta>0, \qquad \gamma>\beta, \qquad |x|<1.$

4. Expand the solution in Ex. 3 into the series, called hypergeometric,

$$B(\beta,\gamma-\beta)\left[1+\frac{\alpha\beta}{1\cdot\gamma}x+\frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot 2\gamma(\gamma+1)}x^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot 2\cdot 3\gamma(\gamma+1)(\gamma+2)}x^3+\cdots\right].$$

5. Establish these results for Bessel's J-functions:

(α) $\displaystyle J_n(x) = \frac{x^n}{2^n\sqrt{\pi}\Gamma(n+\frac{1}{2})}\int_0^\pi \sin^{2n}\phi\cos(x\cos\phi)\,d\phi, \qquad n>-\tfrac{1}{2}.$

(β) $\displaystyle J_n(x) = \frac{1}{\pi}\frac{x^n}{3\cdot 5\cdots(2n-1)}\int_0^\pi \sin^{2n}\phi\cos(x\cos\phi)\,d\phi, \qquad n=0,1,2,3\cdots.$

6. Show $\frac{1}{\pi}\int_0^\pi \cos(n\phi - x\sin\phi)\,d\phi$ satisfies

$$y'' + \frac{y'}{x} + \left(1 - \frac{n^2}{x^2}\right)y = \frac{\sin n\pi}{\pi}\left(\frac{1}{x} - \frac{n}{x^2}\right).$$

7. Find the equation of the second order satisfied by $\int_0^1 (1-t^2)^{n-\frac{1}{2}}\sin xtdt$.

8. Show $J_0(2x) = 1 - x^2 + \frac{x^4}{(2!)^2} - \frac{x^6}{(3!)^2} + \frac{x^8}{(4!)^2} - \frac{x^{10}}{(5!)^2} + \cdots$.

9. Compute $J_0(1) = 0.7652$; $J_0(2) = 0.2239$; $J_0(2.405) = 0.0000$.

10. Prove, from the integrals, $J_0'(x) = -J_1(x)$ and $[x^{-n}J_n]' = -x^{-n}J_{n+1}$.

11. Show that four terms in the asymptotic expansion of $P + iQ$ when $n = 0$ give the best result when $x = 2$ and that the error may be about 0.002.

12. From the asymptotic expansions compute $J_0(3)$ as accurately as may be.

13. Show that for large values of x the solutions of $J_n(x) = 0$ are nearly of the form $k\pi - \frac{1}{4}\pi + \frac{1}{2}n\pi$ and the solutions of $K_n(x) = 0$ of the form $k\pi + \frac{1}{4}\pi + \frac{1}{2}n\pi$.

14. Sketch the graphs of $y = J_0(x)$ and $y = J_1(x)$ by using the series of ascending powers for small values and the asymptotic expressions for large values of x.

15. From $J_0(x) = \frac{1}{\pi}\int_0^\pi \cos(x\cos\phi)\,d\phi$ show $\int_0^\infty e^{-ax}J_0(bx)\,dx = \frac{1}{\sqrt{a^2+b^2}}$.

16. Show $\int_0^\infty e^{-ax}J_0(x)\,dx$ converges uniformly when $a \geqq 0$.

17. Evaluate the following integrals: (α) $\int_0^\infty J_0(bx)\,dx = b^{-1}$,

(β) $\int_0^\infty \sin ax J_0(bx)\frac{dx}{x} = \frac{\pi}{2}$ or $\sin^{-1}\frac{a}{b}$ as $a > b > 0$ or $b > a > 0$,

(γ) $\int_0^\infty \sin ax J_0(bx)\,dx = \frac{1}{\sqrt{a^2-b^2}}$ or 0 as $a^2 > b^2$ or $b^2 > a^2$,

(δ) $\int_0^\infty \cos ax\, J_0(bx)\,dx = \frac{1}{\sqrt{b^2-a^2}}$ or 0 as $b^2 > a^2$ or $a^2 > b^2$.

18. If $u = \sqrt{x}J_n(ax)$, show $\frac{d^2u}{dx^2} + \left(a^2 - \frac{n^2-\frac{1}{4}}{x^2}\right)u = 0$. If $v = \sqrt{x}J_n(bx)$,

$$\left[v\frac{du}{dx} - u\frac{dv}{dx}\right]_0^1 = (b^2 - a^2)\int_0^1 xJ_n(ax)J_n(bx)\,dx.$$

19. With the aid of Ex. 18 establish the relations:

(α) $bJ_n(a)J_{n-1}(b) - aJ_n(b)J_{n-1}(a) = (b^2 - a^2)\int_0^1 xJ_n(ax)J_n(bx)\,dx$.

(β) $aJ_1(a) = a^2\int_0^1 xJ_0(ax)\,dx = \int_0^a xJ_0(x)\,dx$.

(γ) $J_n(a)J_{n+1}(a) + a[J_n(a)J_{n-1}'(a) - J_n'(a)J_{n-1}(a)] = 2a\int_0^1 x[J_n(ax)]^2dx$.

20. Show $J_0(x) = \frac{2}{\pi}\int_1^\infty \frac{\sin xtdt}{\sqrt{t^2-1}}$, $\quad K_0(x) = \frac{2}{\pi}\int_1^\infty \frac{\cos xtdt}{\sqrt{t^2-1}}$.

CHAPTER XV

THE CALCULUS OF VARIATIONS

155. The treatment of the simplest case. The integral

$$I = \int_{C\,A}^{B} F(x, y, y')\,dx = \int_{C\,A}^{B} \Phi(x, y, dx, dy), \tag{1}$$

where Φ is homogeneous of the first degree in dx and dy, may be evaluated along any curve C between the limits A and B by reduction to an ordinary integral. For if C is given by $y = f(x)$,

$$I = \int_{C\,A}^{B} F(x, y, y')\,dx = \int_{x_0}^{x_1} F(x, f(x), f'(x))\,dx;$$

and if C is given by $x = \phi(t)$, $y = \psi(t)$,

$$I = \int_{C\,A}^{B} \Phi(x, y, dx, dy) = \int_{t_0}^{t_1} \Phi(\phi, \psi, \phi', \psi')\,dt.$$

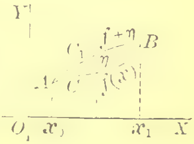
The ordinary line integral (§ 122) is merely the special case in which $\Phi = Pdx + Qdy$ and $F = P + Qy'$. In general the value of I will depend on the path C of integration: *the problem of the calculus of variations is to find that path which will make I a maximum or minimum relative to neighboring paths.*

If a second path C_1 be $y = f(x) + \eta(x)$, where $\eta(x)$ is a small quantity which vanishes at x_0 and x_1, a whole family of paths is given by

$$y = f(x) + \alpha\eta(x), \qquad -1 \leqq \alpha \leqq 1, \qquad \eta(x_0) = \eta(x_1) = 0,$$

and the value of the integral

$$I(\alpha) = \int_{x_0}^{x_1} F(x, f + \alpha\eta, f' + \alpha\eta')\,dx, \tag{1'}$$



taken along the different paths of the family, becomes a function of α; in particular $I(0)$ and $I(1)$ are the values along C and C_1. Under appropriate assumptions as to the continuity of F and its partial derivatives F'_x, F'_y, $F'_{y'}$, the function $I(\alpha)$ will be continuous and have a continuous derivative which may be found by differentiating under the sign (§ 119); then

$$I'(\alpha) = \int_{x_0}^{x_1} [\eta F'_y(x, f + \alpha\eta, f' + \alpha\eta') + \eta' F'_{y'}(x, f + \alpha\eta, f' + \alpha\eta')]\,dx.$$

If the curve C is to give $I(\alpha)$ a maximum or minimum value for all the curves of this family, it is necessary that

$$I'(0)=\int_{x_0}^{x_1}[\eta F'_y(x,\,y,\,y')+\eta' F'_{y'}(x,\,y,\,y')]\,dx=0\,; \tag{2}$$

and if C is to make I a maximum or minimum relative to all neighboring curves, it is necessary that (2) shall hold for any function $\eta(x)$ which is small. It is more usual and more suggestive to write $\eta(x)=\delta y$, and to say that δy is the *variation of* y in passing from the curve C or $y=f(x)$ to the neighboring curve C' or $y=f(x)+\eta(x)$. From the relations

$$y'=f'(x),\qquad y'=f'(x)+\eta'(x),\qquad \delta y'=\eta'(x)=\frac{d}{dx}\delta y,$$

connecting the slope of C with the slope of C_1, it is seen that *the variation of the derivative is the derivative of the variation.* In differential notation this is $d\delta y=\delta dy$, where it should be noted that the sign δ applies to changes which occur on passing from one curve C to another curve C_1, and the sign d applies to changes taking place along a particular curve.

With these notations the condition (2) becomes

$$\int_{x_0}^{x_1}(F'_y\delta y+F'_{y'}\delta y')\,dx=\int_{x_0}^{x_1}\delta F dx=0, \tag{3}$$

where δF is computed from F, δy, $\delta y'$ by the same rule as the differential dF is computed from F and the differentials of the variables which it contains. The condition (3) is not sufficient to distinguish between a maximum and a minimum or to insure the existence of either; neither is the condition $g'(x)=0$ in elementary calculus sufficient to answer these questions relative to a function $g(x)$; in both cases additional conditions are required (§ 9). It should be remembered, however, that these additional conditions were seldom actually applied in discussing maxima and minima of $g(x)$ in practical problems, because in such cases the distinction between the two was usually obvious; so in this case the discussion of sufficient conditions will be omitted altogether, as in §§ 58 and 61, and (3) alone will be applied.

An integration by parts will convert (3) into a differential equation of the second order. In fact

$$\int_{x_0}^{x_1}F'_{y'}\delta y' dx=\int_{x_0}^{x_1}F'_{y'}\frac{d}{dx}\delta y dx=\Big[F'_{y'}\delta y\Big]_{x_0}^{x_1}-\int_{x_0}^{x_1}\delta y\frac{d}{dx}F'_{y'}dx.$$

Hence $$\int_{x_0}^{x_1}(F'_y\delta y+F'_{y'}\delta y')\,dx=\int_{x_0}^{x_1}\left(F'_y-\frac{d}{dx}F'_{y'}\right)\delta y dx=0, \tag{3'}$$

since the assumption that $\delta y = \eta(x)$ vanishes at x_0 and x_1 causes the integrated term $[F'_{y'}\delta y]$ to drop out. Then

$$F'_y - \frac{d}{dx}F'_{y'} = \frac{\partial F}{\partial y} - \frac{\partial^2 F}{\partial x \partial y'} - \frac{\partial^2 F}{\partial y \partial y'}y' - \frac{\partial^2 F}{\partial y'^2}y'' = 0. \tag{4}$$

For it must be remembered that the function $\delta y = \eta(x)$ is *any* function that is small, and if $F'_y - \frac{d}{dx}F'_{y'}$ in (3′) did not vanish at every point of the interval $x_0 \leqq x \leqq x_1$, the arbitrary function δy could be chosen to agree with it in sign, so that the integral of the product would necessarily be positive instead of zero as the condition demands.

156. *The method of rendering an integral* (1) *a minimum or maximum is therefore to set up the differential equation* (4) *of the second order and solve it.* The solution will contain two arbitrary constants of integration which may be so determined that one particular solution shall pass through the points A and B, which are the initial and final points of the path C of integration. In this way a path C which connects A and B and which satisfies (4) is found: under ordinary conditions the integral will then be either a maximum or minimum. An example follows.

Let it be required to render $I = \int_{x_0}^{x_1}\frac{1}{y}\sqrt{1+y'^2}dx$ a maximum or minimum.

$$F(x, y, y') = \frac{1}{y}\sqrt{1+y'^2}, \qquad \frac{\partial F}{\partial y} = -\frac{1}{y^2}\sqrt{1+y'^2}, \qquad \frac{\partial F}{\partial y'} = \frac{y'}{y}\frac{1}{\sqrt{1+y'^2}}.$$

Hence $-\frac{1}{y^2}\sqrt{1+y'^2} + \frac{y'}{y^2}\frac{1}{\sqrt{1+y'^2}}y' - \frac{1}{y}\frac{1}{(1+y'^2)^{\frac{3}{2}}}y'' = 0$ or $yy'' + y'^2 + 1 = 0$

is the desired equation (4). It is exact and the integration is immediate.

$$(yy')' + 1 = 0, \qquad yy' + x = c_1, \qquad y^2 + (x - c_1)^2 = c_2.$$

The curves are circles with their centers on the x-axis. From this fact it is easy by a geometrical construction to determine the curve which passes through two given points $A(x_0, y_0)$ and $B(x_1, y_1)$; the analytical determination is not difficult. The two points A and B must lie on the same side of the x-axis or the integral I will not converge and the problem will have no meaning. The question of whether a maximum or a minimum has been determined may be settled by taking a curve C_1 which lies under the circular arc from A to B and yet has the same length. The integrand is of the form ds/y and the integral along C_1 is greater than along the circle C if y is positive, but less if y is negative. It therefore appears that the integral is rendered a minimum if A and B are above the axis, but a maximum if they are below.

For many problems it is more convenient not to make the choice of x or y as independent variable in the first place, but to operate symmetrically with both variables upon the second form of (1). Suppose that the integral of the variation of Φ be set equal to zero, as in (3).

$$\int_A^B \delta\Phi = \int_A^B [\Phi'_x \delta x + \Phi'_y \delta y + \Phi'_{dx}\delta dx + \Phi'_{dy}\delta dy] = 0.$$

Let the rules $\delta dx = d\delta x$ and $\delta dy = d\delta y$ be applied and let the terms which contain $d\delta x$ and $d\delta y$ be integrated by parts as before.

$$\int_A^B \delta\Phi = \int_A^B [(\Phi'_x - d\Phi'_{dx})\delta x + (\Phi'_y - d\Phi'_{dy})\delta y] + \Big[\Phi'_{dx}\delta x + \Phi'_{dy}\delta y\Big]_A^B = 0.$$

As A and B are fixed points, the integrated term disappears. As the variations δx and δy may be arbitrary, reasoning as above gives

$$\Phi'_x - d\Phi'_{dx} = 0, \qquad \Phi'_y - d\Phi'_{dy} = 0. \tag{4'}$$

If these two equations can be shown to be essentially identical and to reduce to the condition (4) previously obtained, the justification of the second method will be complete and either of (4′) may be used to determine the solution of the problem.

Now the identity $\Phi(x, y, dx, dy) = F(x, y, dy/dx)dx$ gives, on differentiation,

$$\Phi'_x = F'_x dx, \qquad \Phi'_y = F'_y dx, \qquad \Phi'_{dy} = F'_{y'}, \qquad \Phi'_{dx} = -F'_{y'}\frac{dy}{dx} + F$$

by the ordinary rules for partial derivatives. Substitution in each of (4′) gives

$$\Phi'_y - d\Phi'_{dy} = F'_y dx - dF'_{y'} = \left(F'_y - \frac{d}{dx}F'_{y'}\right)dx = 0.$$

$$\begin{aligned}\Phi'_x - d\Phi'_{dx} &= F'_x dx - d(F - F'_{y'}y') = F'_x dx - dF + F'_{y'}dy' + y'dF'_{y'}\\ &= F'_x dx - F'_x dx - F'_y dy - F'_{y'}dy' + F'_{y'}dy' + y'dF'_{y'}\\ &= -F'_y dy + y'dF'_{y'} = -\left(F'_y - \frac{d}{dx}F'_{y'}\right)dy = 0.\end{aligned}$$

Hence each of (4′) reduces to the original condition (4), as was to be proved.

Suppose this method be applied to $\int \frac{ds}{y} = \int \frac{\sqrt{dx^2 + dy^2}}{y}$. Then

$$\int \delta\frac{ds}{y} = \int \delta\frac{\sqrt{dx^2 + dy^2}}{y} = \int\left[\frac{dx\delta dx + dy\delta dy}{yds} - \frac{ds}{y^2}\delta y\right]$$

$$= -\int\left[d\frac{dx}{yds}\delta x + \left(d\frac{dy}{yds} + \frac{ds}{y^2}\right)\delta y\right],$$

where the transformation has been integration by parts, including the discarding of the integrated term which vanishes at the limits. The two equations are

$$d\frac{dx}{yds} = 0, \qquad d\frac{dy}{yds} + \frac{ds}{y^2} = 0; \quad \text{and} \quad \frac{dx}{yds} = \frac{1}{r_1}$$

is the obvious first integral of the first. The integration may then be completed to find the circles as before. The integration of the second equation would not be so simple. In some instances *the advantage of the choice of one of the two equations offered by this method of direct operation is marked.*

EXERCISES

1. *The shortest distance.* Treat $\int (1 + y'^2)^{\frac{1}{2}} dx$ for a minimum.

2. Treat $\int \sqrt{dr^2 + r^2 d\phi^2}$ for a minimum in polar coördinates.

3. *The brachistochrone.* If a particle falls along any curve from A to B, the velocity acquired at a distance h below A is $v = \sqrt{2gh}$ regardless of the path followed. Hence the time spent in passing from A to B is $T = \int ds/v$. The path of quickest descent from A to B is called the brachistochrone. Show that the curve is a cycloid. Take the origin at A.

4. The minimum surface of revolution is found by revolving a catenary.

5. The curve of constant density which joins two points of the plane and has a minimum moment of inertia with respect to the origin is $c_1 r^3 = \sec(3\phi + c_2)$. Note that the two points must subtend an angle of less than 60° at the origin.

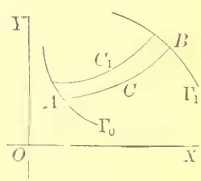
6. Upon the sphere the minimum line is the great circle (polar coördinates).

7. Upon the circular cylinder the minimum line is the helix.

8. Find the minimum line on the cone of revolution.

9. Minimize the integral $\int \left[\frac{1}{2} m \left(\frac{dx}{dt}\right)^2 + \frac{1}{2} n^2 x^2\right] dt.$

157. Variable limits and constrained minima. This second method of operation has also the advantage that it suggests the solution of *the problem of making an integral between variable end-points a maximum or minimum.* Thus suppose that the curve C which shall join some point A of one curve Γ_0 to some point B of another curve Γ_1, and which shall make a given integral a minimum or maximum, is desired. In the first place C must satisfy the condition (4) or (4′) for fixed end-points because C will not give a maximum or minimum value as compared with all other curves unless it does as compared merely with all other curves which join its end-points. There must, however, be additional conditions which shall serve to determine the points A and B which C connects. These conditions are precisely that *the integrated terms,*



$$\left[\Phi'_{dx}\delta x + \Phi'_{dy}\delta y\right]_A^B = 0, \qquad \text{for } A \text{ and for } B, \tag{5}$$

which vanish identically when the end-points are fixed, *shall vanish at each point A or B* provided δx and δy are interpreted as differentials along the curves Γ_0 and Γ_1.

For example, in the case of $\int \frac{ds}{y} = \int \frac{\sqrt{dx^2 + dy^2}}{y}$ treated above, the integrated terms, which were discarded, and the resulting conditions are

$$\left[\frac{dx\delta x}{yds} + \frac{dy\delta y}{yds}\right]_A^B, \qquad \frac{dx\delta x + dy\delta y}{yds}\Bigg]^B = 0, \qquad \frac{dx\delta x + dy\delta y}{yds}\Bigg]_A = 0.$$

Here dx and dy are differentials along the circle C and δx and δy are to be interpreted as differentials along the curves Γ_0 and Γ_1 which respectively pass through A and B. The conditions therefore show that the tangents to C and Γ_0 at A are perpendicular, and similarly for C and Γ_1 at B. In other words the curve which renders the integral a minimum and has its extremities on two fixed curves is the circle which has its center on the x-axis and cuts both the curves orthogonally.

To prove the rule for finding the conditions at the end points it will be sufficient to prove it for one variable point. Let the equations

$$C : x = \phi(t), \qquad y = \psi(t), \qquad C_1 : x = \phi(t) + \zeta(t), \qquad y = \psi(t) + \eta(t),$$

$$\zeta(t_0) = \eta(t_0) = 0, \qquad \zeta(t_1) = a, \qquad \eta(t_1) = b; \qquad \delta x = \zeta(t), \qquad \delta y = \eta(t),$$

determine C and C_1 with the common initial point A and different terminal points B and B' upon Γ_1. As parametric equations of Γ_1, take

$$x = x_B + al(s), \qquad y = y_B + bm(s); \qquad \frac{\delta x}{\delta s} = al'(s), \qquad \frac{\delta y}{\delta s} = bm'(s),$$

where s represents the arc along Γ_1 measured from B, and the functions $l(s)$ and $m(s)$ vary from 0 at B to 1 at B'. Next form the family

$$x = \phi(t) + l(s)\zeta(t), \qquad y = \psi(t) + m(s)\eta(t), \qquad x' = \phi' + l\zeta', \qquad y' = \psi' + m\eta',$$

which all pass through A for $t = t_0$ and which for $t = t_1$ describe the curve Γ_1. Consider

$$g(s) = \int_{t_0}^{t_1} \Phi(x + l(s)\zeta,\ y + m(s)\eta,\ x' + l\zeta',\ y' + m\eta')\,dt, \tag{6}$$

which is the integral taken from A to Γ_1 along the curves of the family, where x, y, x', y' are on the curve C corresponding to $s = 0$. Differentiate. Then

$$g'(s) = \int_{t_0}^{t_1} \left[l'(s)\zeta\Phi'_x + m'(s)\eta\Phi'_y + l'(s)\zeta'\Phi'_{x'} + m'(s)\eta'\Phi'_{y'}\right]dt,$$

where the accents mean differentiation with regard to s when upon g, l, or m, but with regard to t when on x or y, and partial differentiation when on Φ, and where the argument of Φ is as in (6). Now if $g(s)$ has a maximum or minimum when $s = 0$, then

$$g'(0) = \int_{t_0}^{t_1} \left[l'(0)\zeta\Phi'_x(x, y, x', y') + m'(0)\eta\Phi'_y + l'(0)\zeta'\Phi'_{x'} + m'(0)\eta'\Phi'_{y'}\right]dt = 0:$$

$$\left[l'(0)\zeta\Phi'_{x'} + m'(0)\eta\Phi'_{y'}\right]_{t_0}^{t_1} + \int_{t_0}^{t_1}\left[l'(0)\zeta\left(\Phi'_x - \frac{d}{dt}\Phi'_{x'}\right) + m'(0)\eta\left(\Phi'_y - \frac{d}{dt}\Phi'_{y'}\right)\right]dt = 0.$$

The change is made as usual by integration by parts. Now as

$$\Phi(x, y, x', y')\,dt = \Phi(x, y, dx, dy), \quad \text{so} \quad \Phi'_x\,dt = \Phi'_x, \qquad \Phi'_{x'} = \Phi'_{dx}, \text{ etc.}$$

Hence the parentheses under the integral sign, when multiplied by dt, reduce to (4′) and vanish; they could be seen to vanish also for the reason that ζ and η are arbitrary functions of t except at $t = t_0$ and $t = t_1$, and the integrated term is a constant. There remains the integrated term which must vanish,

$$l'(0)\,\zeta(t_1)\,\Phi'_{x'} + m'(0)\,\eta(t_1)\,\Phi'_{y'} = \left[\frac{\delta x}{\delta s}\Phi'_{x'} + \frac{\delta y}{\delta s}\Phi'_{y'}\right]^{t_1} = \left[\Phi'_{dx}\,\delta x + \Phi'_{dy}\,\delta y\right]^{t_1} = 0.$$

The condition therefore reduces to its appropriate half of (5), provided that, in interpreting it, the quantities δx and δy be regarded not as $a = \zeta(t_1)$ and $b = \eta(t_1)$ but as the differentials along Γ_1 at B.

158. In many cases one integral is to be made a maximum or minimum subject to the condition that another integral shall have a fixed value,

$$I = \int_{x_0}^{x_1} F(x, y, y')\,dx \;\frac{\text{min.}}{\text{max.}}, \qquad J = \int_{x_0}^{x_1} G(x, y, y')dx = \text{const.} \quad (7)$$

For instance a curve of given length might run from A to B, and the form of the curve which would make the area under the curve a maximum or minimum might be desired; to make the area a maximum or minimum without the restriction of constant length of arc would be useless, because by taking a curve which dropped sharply from A, inclosed a large area below the x-axis, and rose sharply to B the area could be made as small as desired. Again the curve in which a chain would hang might be required. The length of the chain being given, the form of the curve is that which will make the potential energy a minimum, that is, will bring the center of gravity lowest. The problems in constrained maxima and minima are called *isoperimetric problems* because it is so frequently the perimeter or length of the curve which is given as constant.

If the method of determining constrained maxima and minima by means of undetermined multipliers be recalled (§§ 58, 61), it will appear that the solution of the isoperimetric problem might reasonably be sought by rendering the integral

$$I + \lambda J = \int_{x_0}^{x_1} [F(x, y, y') + \lambda G(x, y, y')]\,dx \quad (8)$$

a maximum or minimum. The solution of this problem would contain three constants, namely, λ and two constants c_1, c_2 of integration. The constants c_1, c_2 could be determined so that the curve should pass through A and B and the value of λ would still remain to be determined in such a manner that the integral J should have the desired value. This is the method of solution.

To justify the method in the case of fixed end-points, which is the only case that will be considered, the procedure is like that of § 155. Let C be given by $y = f(x)$; consider

$$y = f(x) + \alpha\eta(x) + \beta\zeta(x), \qquad \eta_0 = \eta_1 = \zeta_0 = \zeta_1 = 0,$$

a two-parametered family of curves near to C. Then

$$g(\alpha, \beta) = \int_{x_0}^{x_1} F(x, y + \alpha\eta + \beta\zeta, y' + \alpha\eta' + \beta\zeta')\,dx, \qquad g(0, 0) = I$$

$$h(\alpha, \beta) = \int_{x_0}^{x_1} G(x, y + \alpha\eta + \beta\zeta, y' + \alpha\eta' + \beta\zeta')\,dx = J = \text{const.}$$

would be two functions of the two variables α and β. The conditions for the minimum or maximum of $g(\alpha, \beta)$ at $(0, 0)$ subject to the condition that $h(\alpha, \beta) = \text{const.}$ are required. Hence

$$g'_\alpha(0, 0) + \lambda h'_\alpha(0, 0) = 0, \qquad g'_\beta(0, 0) + \lambda h'_\beta(0, 0) = 0,$$

or

$$\int_{x_0}^{x_1} \eta\,(F'_y + \lambda G'_y) + \eta'(F'_{y'} + \lambda G'_{y'})\,dx = 0,$$

$$\int_{x_0}^{x_1} \zeta\,(F'_y + \lambda G'_y) + \zeta'(F'_{y'} + \lambda G'_{y'})\,dx = 0.$$

By integration by parts either of these equations gives

$$(F + \lambda G)'_y - \frac{d}{dx}(F + \lambda G)'_{y'} = 0\,; \tag{9}$$

the rule is justified, and will be applied to an example.

Required the curve which, when revolved about an axis, will generate a given volume of revolution bounded by the least surface. The integrals are

$$I = 2\pi \int_{x_0}^{x_1} y\,ds, \text{ min.}, \qquad J = \pi \int_{x_0}^{x_1} y^2 dx, \text{ const.}$$

Make

$$\int_{x_0}^{x_1} (y\,ds + \lambda y^2 dx) \text{ min.} \quad \text{or} \quad \int_{x_0}^{x_1} \delta\,(y\,ds + \lambda y^2 dx) = 0.$$

$$\int_{x_0}^{x_1} \delta\,(y\,ds + \lambda y^2 dx) = \int_{x_0}^{x_1} \left[\delta y\,ds + y\,\frac{dx\,\delta dx + dy\,\delta dy}{ds} + 2\lambda y\,\delta y\,dx + \lambda y^2 \delta dx\right] = 0$$

$$= \int_{x_0}^{x_1} \left[\delta x\left(-\lambda d\,(y^2) - d\,\frac{y\,dx}{ds}\right) + \delta y\left(ds - d\,\frac{y\,dy}{ds} + 2\lambda y\,dx\right)\right].$$

Hence

$$\lambda d\,(y^2) + d\,\frac{y\,dx}{ds} = 0 \quad \text{or} \quad ds - d\,\frac{y\,dy}{ds} + 2\lambda y\,dx = 0.$$

The second method of computation has been used and the vanishing integrated terms have been discarded. The first equation is simplest to integrate.

$$\lambda y^2 + y\,\frac{1}{\sqrt{1 + y'^2}} = c_1\lambda, \qquad \pm\frac{\lambda\,(c_1 - y^2)\,dy}{\sqrt{y^2 - \lambda^2(c_1 - y^2)^2}} = dx.$$

The variables are separated, but the integration cannot be executed in terms of elementary functions. If, however, one of the end-points is on the x-axis, the

values x_0, 0, y'_0 or x_1, 0, y'_1 must satisfy the equation and, as no term of the equation can become infinite, c_1 must vanish. The integration may then be performed.

$$\pm\frac{\lambda y\,dy}{\sqrt{1-\lambda^2 y^2}}=dx,\qquad 1-\lambda^2y^2=\lambda^2(x-c_2)^2\quad\text{or}\quad (x-c_2)^2+y^2=\frac{1}{\lambda^2}.$$

In this special case the curve is a circle. The constants c_1 and λ may be determined from the other point (x_1, y_1) through which the curve passes and from the value of $J=v$; the equations will also determine the abscissa x_0 of the point on the axis. It is simpler to suppose $x_0=0$ and leave x_1 to be determined. With this procedure the equations are

$$c_2^2=\frac{1}{\lambda^2},\qquad (x_1-c_2)^2+y_1^2=\frac{1}{\lambda^2},\qquad \frac{v}{\pi}=\frac{x_1}{\lambda^2}-\frac{1}{3}(x_1^3-3c_2x_1^2+3c_2^2x_1),$$

or
$$x_1^3+3y_1^2x_1-\frac{6v}{\pi}=0,\qquad c_2=\frac{x_1^2+y_1^2}{2x_1},$$

and
$$x_1=\pi^{-\frac{1}{3}}\left[\left(3v+\sqrt{9v^2+\pi^2y_1^6}\right)^{\frac{1}{3}}+\left(3v-\sqrt{9v^2+\pi^2y_1^6}\right)^{\frac{1}{3}}\right].$$

EXERCISES

1. Show that (α) the minimum line from one curve to another in the plane is their common normal; (β) if the ends of the catenary which generates the minimum surface of revolution are constrained to lie on two curves, the catenary shall be perpendicular to the curves; (γ) the brachistochrone from a fixed point to a curve is the cycloid which cuts the curve orthogonally.

2. Generalize to show that if the end-points of the curve which makes any integral of the form $\int F(x, y)\,ds$ a maximum or a minimum are variable upon two curves, the solution shall cut the curves orthogonally.

3. Show that if the integrand $\Phi(x, y, dx, dy, x_1)$ depends on the limit x_1, the condition for the limit B becomes $\left[\Phi'_{dx}\delta x+\Phi'_{dy}\delta y+\delta x\int_{x_0}^{x_1}\Phi'_{x_1}\right]^B=0.$

4. Show that the cycloid which is the brachistochrone from a point A, constrained to lie on one curve Γ_0, to another curve Γ_1 must leave Γ_0 at the point A where the tangent to Γ_0 is parallel to the tangent to Γ_1 at the point of arrival.

5. Prove that the curve of given length which generates the minimum surface of revolution is still the catenary.

6. If the area under a curve of given length is to be a maximum or minimum, the curve must be a circular arc connecting the two points.

7. In polar coördinates the sectorial area bounded by a curve of given length is a maximum or minimum when the curve is a circle.

8. A curve of given length generates a maximum or minimum volume of revolution. The elastic curve

$$R=\frac{(1+y'^2)^{\frac{3}{2}}}{y''}=-\frac{\lambda}{2y}\quad\text{or}\quad dx=\frac{(y^2-c_1)\,dy}{\sqrt{\lambda^2-(y^2-c_1)^2}}.$$

9. A chain lies in a central field of force of which the potential per unit mass is $V(r)$. If the constant density of the chain is ρ, show that the form of the curve is

$$\phi + c_2 = \int \frac{dr}{r\left[c_1^2(\rho V + \lambda)^2 r^2 - 1\right]^{\frac{1}{2}}}.$$

10. Discuss the reciprocity of I and J, that is, the questions of making I a maximum or minimum when J is fixed, and of making J a minimum or maximum when I is fixed.

11. A solid of revolution of given mass and uniform density exerts a maximum attraction on a point at its axis. *Ans.* $2\lambda(x^2 + y^2)^{\frac{3}{2}} + x = 0$, if the point is at the origin.

159. Some generalizations. Suppose that an integral

$$I = \int_A^B F(x, y, y', z, z', \cdots)\, dx = \int_A^B \Phi(x, dx, y, dy, z, dz, \cdots) \qquad (10)$$

(of which the integrand contains two or more dependent variables $y, z, \cdots$ and their derivatives $y', z', \cdots$ with respect to the independent variable x, or in the symmetrical form contains three or more variables and their differentials) were to be made a maximum or minimum. In case there is only one additional variable, the problem still has a geometric interpretation, namely, to find

$$y = f(x), \qquad z = g(x), \quad \text{or} \quad x = \phi(t), \qquad y = \psi(t), \qquad z = \chi(t),$$

a curve in space, which will make the value of the integral greater or less than all neighboring curves. A slight modification of the previous reasoning will show that necessary conditions are

$$F'_y - \frac{d}{dx} F'_{y'} = 0 \quad \text{and} \quad F'_z - \frac{d}{dx} F'_{z'} = 0 \qquad (11)$$

$$\text{or} \qquad \Phi'_x - d\Phi'_{dx} = 0, \qquad \Phi'_y - d\Phi'_{dy} = 0, \qquad \Phi'_z - d\Phi'_{dz} = 0,$$

where of the last three conditions only two are independent. Each of (11) is a differential equation of the second order, and the solution of the two simultaneous equations will be a family of curves in space dependent on four arbitrary constants of integration which may be so determined that one curve of the family shall pass through the end-points A and B.

Instead of following the previous method to establish these facts, an older and perhaps less accurate method will be used. Let the varied values of y, z, y', z', be denoted by

$$y + \delta y, \quad z + \delta z, \quad y' + \delta y', \quad z' + \delta z', \quad \delta y' = (\delta y)', \quad \delta z' = (\delta z)'.$$

The difference between the integral along the two curves is

$$\Delta I=\int_{x_0}^{x_1}[F(x,\,y+\delta y,\,y'+\delta y',\,z+\delta z,\,z'+\delta z')-F(x,\,y,\,y',\,z,\,z')]\,dx$$

$$=\int_{x_0}^{x_1}\Delta F dx=\int_{x_0}^{x_1}(F'_y\delta y+F'_{y'}\delta y'+F'_z\delta z+F'_{z'}\delta z')\,dx+\cdots,$$

where F has been expanded by Taylor's Formula* for the four variables y, y', z, z' which are varied, and "$+\cdots$" refers to the remainder or the subsequent terms in the development which contain the higher powers of $\delta y,\ \delta y',\ \delta z,\ \delta z'$.

For sufficiently small values of the variations the terms of higher order may be neglected. Then if ΔI is to be either positive or negative for all small variations, the terms of the first order which change in sign when the signs of the variations are reversed must vanish and the condition becomes

$$\int_{x_0}^{x_1}(F'_y\delta y+F'_{y'}\delta y'+F'_z\delta z+F'_{z'}\delta z')\,dx=\int_{x_0}^{x_1}\delta F dx=0. \tag{12}$$

Integrate by parts and discard the integrated terms. Then

$$\int_{x_0}^{x_1}\left[\left(F'_y-\frac{d}{dx}F'_{y'}\right)\delta y+\left(F'_z-\frac{d}{dx}F'_{z'}\right)\delta z\right]=0. \tag{13}$$

* In the simpler case of § 155 this formal development would run as

$$\Delta I=\int_{x_0}^{x_1}(F'_y\delta y+F'_{y'}\delta y')\,dx+\frac{1}{2!}\int_{x_0}^{x_1}(F''_{yy}\delta y^2+2F''_{yy'}\delta y\delta y'+F''_{y'y'}\delta y'^2)dx+\text{higher terms},$$

and with the expansion $\Delta I=\delta I+\frac{1}{2!}\delta^2 I+\frac{1}{3!}\delta^3 I+\cdots$ it would appear that

$$\delta I=\int_{x_0}^{x_1}(F'_y\delta y+F'_{y'}\delta y')dx,\qquad \delta^2 I=\int_{x_0}^{x_1}(F''_{yy}\delta y^2+2F''_{yy'}\delta y\delta y'+F''_{y'y'}\delta y'^2)dx,$$

$$\delta^3 I=\int_{x_0}^{x_1}(F'''_{y^3}\delta y^3+3F'''_{y^2y'}\delta y^2\delta y'+3F'''_{yy'^2}\delta y\delta y'^2+F'''_{y'^3}\delta y'^3)dx,\ \cdots.$$

The terms $\delta I,\ \delta^2 I,\ \delta^3 I,\ \cdots$ are called the *first, second, third,* $\cdots$ *variations* of the integral I in the case of fixed limits. The condition for a maximum or minimum then becomes $\delta I=0$, just as $dg=0$ is the condition in the case of $g(x)$. In the case of variable limits there are some modifications appropriate to the limits. This method of procedure suggests the reason that $\delta x,\ \delta y$ are frequently to be treated exactly as differentials. It also suggests that $\delta^2 I>0$ and $\delta^2 I<0$ would be criteria for distinguishing between maxima and minima. The same results can be had by differentiating (1′) repeatedly under the sign and expanding $I(\alpha)$ into series; in fact, $\delta I=I'(0),\ \delta^2 I=I''(0),\ \cdots$. No emphasis has been laid in the text on the suggestive relations $\delta I=\int\delta F dx$ for fixed limits or $\delta I=\int\delta\Phi$ for variable limits (variable in x, y, but not in t) because only the most elementary results were desired, and the treatment given has some advantages as to modernity.

As δy and δz are arbitrary, either may in particular be taken equal to 0 while the other is assigned the same sign as its coefficient in the parenthesis; and hence the integral would not vanish unless that coefficient vanished. Hence the conditions (11) are derived, and it is seen that there would be precisely similar conditions, one for each variable $y, z, \cdots$, no matter how many variables might occur in the integrand.

Without going at all into the matter of proof it will be stated as a fact that the condition for the maximum or minimum of

$$\int \Phi(x, dx, y, dy, z, dz, \ldots) \quad \text{is} \quad \int \delta\Phi = 0,$$

which may be transformed into the set of differential equations

$$\Phi'_x - d\Phi'_{dx} = 0, \qquad \Phi'_y - d\Phi'_{dy} = 0, \qquad \Phi'_z - d\Phi'_{dz} = 0, \qquad \cdots,$$

of which any one may be discarded as dependent on the rest; and

$$\Phi'_{dx}\delta x + \Phi'_{dy}\delta y + \Phi'_{dz}\delta z + \cdots = 0, \qquad \text{at } A \text{ and at } B,$$

where the variations are to be interpreted as differentials along the loci upon which A and B are constrained to lie.

It frequently happens that the variables in the integrand of an integral which is to be made a maximum or minimum are connected by an equation. For instance

$$\int \Phi(x, dx, y, dy, z, dz) \text{ min.}, \qquad S(x, y, z) = 0. \tag{14}$$

It is possible to eliminate one of the variables and its differential by means of $S = 0$ and proceed as before; but it is usually better to introduce an undetermined multiplier (§§ 58, 61). From

$$S(x, y, z) = 0 \quad \text{follows} \quad S'_x\delta x + S'_y\delta y + S'_z\delta z = 0$$

if the variations be treated as differentials. Hence if

$$\int [(\Phi'_x - d\Phi'_{dx})\,\delta x + (\Phi'_y - d\Phi'_{dy})\,\delta y + (\Phi'_z - d\Phi'_{dz})\,\delta z] = 0,$$

$$\int [(\Phi'_x - d\Phi'_{dx} + \lambda S'_x)\,\delta x + (\Phi'_y - d\Phi'_{dy} + \lambda S'_y)\,\delta y \\ + (\Phi'_z - d\Phi'_{dz} + \lambda S'_z)\,\delta z] = 0$$

no matter what the value of λ. Let the value of λ be so chosen as to annul the coefficient of δz. Then as the two remaining variations are independent, the same reasoning as above will cause the coefficients of δx and δy to vanish and

$$\Phi'_x - d\Phi'_{dx} + \lambda S'_x = 0, \quad \Phi'_y - d\Phi'_{dy} + \lambda S'_y = 0, \quad \Phi'_z - d\Phi'_{dz} + \lambda S'_z = 0 \tag{15}$$

will hold. These equations, taken with $S = 0$, will determine y and z as functions of x and also incidentally will fix λ.

Consider the problem of determining *the shortest lines upon a surface* $S(x, y, z) = 0$. These lines are called *the geodesics*. Then

$$\int \delta ds = 0 = \frac{dx\delta x + dy\delta y + dz\delta z}{ds}\bigg| - \int \left[d\frac{dx}{ds}\delta x + d\frac{dy}{ds}\delta y + d\frac{dz}{ds}\delta z \right], \quad (16)$$

$$\int \left(d\frac{dx}{ds} + \lambda S'_x \right)\delta x + \left(d\frac{dy}{ds} + \lambda S'_y \right)\delta y + \left(d\frac{dz}{ds} + \lambda S'_z \right)\delta z = 0,$$

$$d\frac{dx}{ds} + \lambda S'_x = d\frac{dy}{ds} + \lambda S'_y = d\frac{dz}{ds} + \lambda S'_z = 0, \quad \text{and} \quad \frac{d\frac{dx}{ds}}{S'_x} = \frac{d\frac{dy}{ds}}{S'_y} = \frac{d\frac{dz}{ds}}{S'_z}.$$

In the last set of equations λ has been eliminated and the equations, taken with $S = 0$, may be regarded as *the differential equations of the geodesics*. The denominators are proportional to the direction cosines of the normal to the surface, and the numerators are the components of the differential of the unit tangent to the curve and are therefore proportional to the direction cosines of the normal to the curve in its osculating plane. Hence it appears that *the osculating plane of a geodesic curve contains the normal to the surface*.

The integrated terms $dx\delta x + dy\delta y + dz\delta z = 0$ show that the least geodesic which connects two curves on the surface will cut both curves orthogonally. These terms will also suffice to prove a number of interesting theorems which establish an analogy between geodesics on a surface and straight lines in a plane. For instance: The locus of points whose geodesic distance from a fixed point is constant (a geodesic circle) cuts the geodesic lines orthogonally. To see this write

$$\int_O^P ds = \text{const.}, \quad \Delta\int_O^P ds = 0, \quad \delta\int_O^P ds = 0, \quad \int_O^P \delta ds = 0 = dx\delta x + dy\delta y + dz\delta z \Big|^P.$$

The integral in (16) drops out because taken along a geodesic. This final equality establishes the perpendicularity of the lines. The fact also follows from the statement that the geodesic circle and its center can be regarded as two curves between which the shortest distance is the distance measured along any of the geodesic radii, and that the radii must therefore be perpendicular to the curve.

160. The most fundamental and important single theorem of mathematical physics is Hamilton's Principle, which is expressed by means of the calculus of variations and affords a necessary and sufficient condition for studying the elements of this subject. Let T be the kinetic energy of any dynamical system. Let X_i, Y_i, Z_i be the forces which act at any point x_i, y_i, z_i of the system, and let $\delta x_i, \delta y_i, \delta z_i$ represent displacements of that point. Then the work is

$$\delta W = \sum (X_i\delta x_i + Y_i\delta y_i + Z_i\delta z_i).$$

Hamilton's Principle states that the *time integral*

$$\int_{t_0}^{t_1}(\delta T+\delta W)\,dt=\int_{t_0}^{t_1}\left[\delta T+\sum(X\delta x+Y\delta y+Z\delta z)\right]dt=0 \quad (17)$$

vanishes for the actual motion of the system. If in particular there is a potential function V, then $\delta W=-\delta V$ and

$$\int_{t_0}^{t_1}\delta(T-V)\,dt=\delta\int_{t_0}^{t_1}(T-V)\,dt=0, \qquad (17')$$

and *the time integral of the difference between the kinetic and potential energies is a maximum or minimum for the actual motion of the system* as compared with any neighboring motion.

Suppose that the position of a system can be expressed by means of n independent variables or coördinates $q_1, q_2, \cdots, q_n$. Let the kinetic energy be expressed as

$$T=\sum\tfrac{1}{2}m_iv_i^2=\int\tfrac{1}{2}v^2dm=T(q_1, q_2, \cdots, q_n, \dot q_1, \dot q_2, \cdots, \dot q_n),$$

a function of the coördinates and their derivatives with respect to the time. Let the work done by displacing the single coördinate q_r be $\delta W=Q_r\delta q_r$, so that the total work, in view of the independence of the coördinates, is $Q_1\delta q_1+Q_2dq_2+\cdots+Q_ndq_n$. Then

$$0=\int_{t_0}^{t_1}(\delta T+\delta W)\,dt=\int_{t_0}^{t_1}(T'_{q_1}\delta q_1+T'_{q_2}\delta q_2+\cdots+T'_{q_n}\delta q_n+T'_{\dot q_1}\delta\dot q_1+T'_{\dot q_2}\delta\dot q_2$$
$$+\cdots+T'_{\dot q_n}\delta\dot q_n+Q_1\delta q_1+Q_2\delta q_2+\cdots+Q_n\delta q_n)\,dt.$$

Perform the usual integration by parts and discard the integrated terms which vanish at the limits $t=t_0$ and $t=t_1$. Then

$$0=\int_{t_0}^{t_1}\left[\left(T'_{q_1}+Q_1-\frac{d}{dt}T'_{\dot q_1}\right)\delta q_1+\left(T'_{q_2}+Q_2-\frac{d}{dt}T'_{\dot q_2}\right)\delta q_2\right.$$
$$\left.+\cdots+\left(T'_{q_n}+Q_n-\frac{d}{dt}T'_{\dot q_n}\right)\delta q_n\right]dt.$$

In view of the independence of the variations $\delta q_1, \delta q_2, \cdots, \delta q_n$,

$$\frac{d}{dt}\frac{\partial T}{\partial\dot q_1}-\frac{\partial T}{\partial q_1}=Q_1,\qquad \frac{d}{dt}\frac{\partial T}{\partial\dot q_2}-\frac{\partial T}{\partial q_2}=Q_2,\qquad\cdots,\qquad \frac{d}{dt}\frac{\partial T}{\partial\dot q_n}-\frac{\partial T}{\partial q_n}=Q_n. \quad (18)$$

These are the *Lagrangian equations* for the motion of a dynamical system.* If there is a potential function $V(q_1, q_2, \cdots, q_n)$, then by definition

$$Q_1=-\frac{\partial V}{\partial q_1},\quad Q_2=-\frac{\partial V}{\partial q_2},\quad\cdots,\quad Q_n=-\frac{\partial V}{\partial q_n},\quad \frac{\partial V}{\partial\dot q_1}=\frac{\partial V}{\partial\dot q_2}=\cdots=\frac{\partial V}{\partial\dot q_n}=0.$$

Hence $\quad \dfrac{d}{dt}\dfrac{\partial L}{\partial\dot q_1}-\dfrac{\partial L}{\partial q_1}=0,\quad \dfrac{d}{dt}\dfrac{\partial L}{\partial\dot q_2}-\dfrac{\partial L}{\partial q_2}=0,\quad\cdots,\quad \dfrac{d}{dt}\dfrac{\partial L}{\partial\dot q_n}-\dfrac{\partial L}{\partial q_n}=0,\quad L=T-V.$

The equations of motion have been expressed in terms of a single function L, which is the difference between the kinetic energy T and potential function V. By

* Compare Ex. 19, p. 112, for a deduction of (18) by transformation.

comparing the equations with (17′) it is seen that the dynamics of a system which may be specified by n coördinates, and which has a potential function, may be stated as the problem of rendering the integral $\int Ldt$ a maximum or a minimum; both the kinetic energy T and potential function V may contain the time t without changing the results.

For example, let it be required to derive the equations of motion of a lamina lying in a plane and acted upon by any forces in the plane. Select as coördinates the ordinary coördinates (x, y) of the center of gravity and the angle ϕ through which the lamina may turn about its center of gravity. The kinetic energy of the lamina (p. 318) will then be the sum $\frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2$. Now if the lamina be moved a distance δx to the right, the work done by the forces will be $X\delta x$, where X denotes the sum of all the components of force along the x-axis no matter at what points they act. In like manner $Y\delta y$ will be the work for a displacement δy. Suppose next that the lamina is rotated about its center of gravity through the angle $\delta\phi$; the actual displacement of any point is $r\delta\phi$ where r is its distance from the center of gravity. The work of any force will then be $Rrd\phi$ where R is the component of the force perpendicular to the radius r; but $Rr = \Phi$ is the moment of the force about the center of gravity. Hence

$$T = \tfrac{1}{2}M(\dot{x}^2 + \dot{y}^2) + \tfrac{1}{2}I\dot{\phi}^2, \qquad \delta W = X\delta x + Y\delta y + \Phi\delta\phi$$

and
$$M\frac{d^2x}{dt^2} = X, \qquad M\frac{d^2y}{dt^2} = Y, \qquad I\frac{d^2\phi}{dt^2} = \Phi,$$

by substitution in (18), are the desired equations, where X and Y are the *total* components along the axis and Φ is the *total* moment about the center of gravity.

A particle glides without friction on the interior of an inverted cone of revolution; determine the motion. Choose the distance r of the particle from the vertex and the meridional angle ϕ as the two coördinates. If l be the sine of the angle between the axis of the cone and the elements, then $ds^2 = dr^2 + r^2l^2d\phi^2$ and $v^2 = \dot{r}^2 + r^2l^2\dot{\phi}^2$. The pressure of the cone against the particle does no work; it is normal to the motion. For a change $\delta\phi$ gravity does no work; for a change δr it does work to the amount $-mg\sqrt{1-l^2}\delta r$. Hence

$$T = \tfrac{1}{2}m(\dot{r}^2 + r^2l^2\dot{\phi}^2), \qquad \delta W = -mg\sqrt{1-l^2}\delta r \quad \text{or} \quad V = mg\sqrt{1-l^2}r.$$

Then
$$\frac{d^2r}{dt^2} - rl^2\left(\frac{d\phi}{dt}\right)^2 = -g\sqrt{1-l^2}, \qquad \frac{d}{dt}\left(r^2l^2\frac{d\phi}{dt}\right) = 0 \quad \text{or} \quad r^2\frac{d\phi}{dt} = C.$$

The remaining integrations cannot all be effected in terms of elementary functions.

161. Suppose the double integral

$$I = \iint F(x, y, z, p, q)dxdy, \qquad p = \frac{\partial z}{\partial x}, \qquad q = \frac{\partial z}{\partial y}, \tag{19}$$

extended over a certain area of the xy-plane were to be made a maximum or minimum by a surface $z = z(x, y)$, which shall pass through a given curve upon the cylinder which stands upon the bounding curve of the area. This problem is analogous to the problem of § 155 with

fixed limits; the procedure for finding the partial differential equation which z shall satisfy is also analogous. Set

$$\iint \delta F dxdy = \iint (F'_z \delta z + F'_p \delta p + F'_q \delta q)\, dxdy = 0.$$

Write $\delta p = \frac{\partial \delta z}{\partial x} \cdot \delta q = \frac{\partial \delta z}{\partial y}$ and integrate by parts.

$$\iint F'_p \frac{\partial \delta z}{\partial x} dxdy = \int F'_p \delta z \Big|_A^B dy - \iint \frac{dF'_p}{dx} \delta z dxdy.$$

The limits A and B for which the first term is taken are points upon the bounding contour of the area, and $\delta z = 0$ for A and B by virtue of the assumption that the surface is to pass through a fixed curve above that contour. The integration of the term in δq is similar. Hence the condition becomes

$$\iint \delta F dxdy = \iint \left(F'_z - \frac{d}{dx}\frac{\partial F}{\partial p} - \frac{d}{dy}\frac{\partial F}{\partial q} \right) \delta z dxdy = 0 \tag{20}$$

or

$$\frac{\partial F}{\partial z} - \frac{d}{dx}\frac{\partial F}{\partial p} - \frac{d}{dy}\frac{\partial F}{\partial q} = 0, \tag{20'}$$

by the familiar reasoning. The total differentiations give

$$F'_z - F''_{xp} - F''_{yq} - F''_{zp} p - F''_{zq} q - F''_{pp} r - 2\, F''_{pq} s - F''_{qq} t = 0.$$

The stock illustration introduced at this point is the minimum surface, that is, the surface which spans a given contour with the least area and which is physically represented by a soap film. The real use, however, of the theory is in connection with Hamilton's Principle. To study the motion of a chain hung up and allowed to vibrate, or of a piano wire stretched between two points, compute the kinetic and potential energies and apply Hamilton's Principle. Is the motion of a vibrating elastic body to be investigated? Apply Hamilton's Principle. And so in electrodynamics. In fact, with the very foundations of mechanics sometimes in doubt owing to modern ideas on electricity, the one refuge of many theorists is Hamilton's Principle. Two problems will be worked in detail to exhibit the method.

Let a uniform chain of density ρ and length l be suspended by one extremity and caused to execute small oscillations in a vertical plane. At any time the shape of the curve is $y = y(x)$, and $y = y(x, t)$ will be taken to represent the shape of the curve at all times. Let $y' = \partial y/\partial x$ and $\dot{y} = \partial y/\partial t$. As the oscillations are small, the chain will rise only slightly and the main part of the kinetic energy will be in the whipping motion from side to side; the assumption $dx = ds$ may be made and the kinetic energy may be taken as

$$T = \int_0^l \frac{1}{2} \rho \left(\frac{\partial y}{\partial t}\right)^2 dx.$$

The potential energy is a little harder to compute, for it is necessary to obtain the slight rise in the center of gravity due to the bending of the chain. Let λ be the shortened length. The position of the center of gravity is

$$\bar{x} = \frac{\int_0^\lambda x\,(1 + \frac{1}{2} y'^2)\,dx}{\int_0^\lambda (1 + \frac{1}{2} y'^2)\,dx} = \frac{\frac{1}{2}\lambda^2 + \int_0^\lambda \frac{1}{2} x y'^2 dx}{\lambda + \int_0^\lambda \frac{1}{2} y'^2 dx} = \frac{1}{2}\lambda - \frac{1}{\lambda}\int_0^\lambda \left(\frac{1}{4}\lambda - \frac{1}{2}x\right) y'^2 dx.$$

Here $ds = \sqrt{1 + y'^2}dx$ has been expanded and terms higher than y'^2 have been omitted.

$$l = \lambda + \int_0^\lambda \frac{1}{2} y'^2 dx, \qquad \frac{1}{2} l - \bar{x} = \frac{1}{\lambda}\int_0^\lambda \frac{1}{2}(\lambda - x)\, y'^2 dx, \qquad V = l\rho g\left(\frac{1}{2} l - \bar{x}\right).$$

Then $$\int_{t_0}^{t_1}(T - V)\,dt = \int_{t_0}^{t_1}\int_0^l \left[\frac{1}{2}\rho\left(\frac{\partial y}{\partial t}\right)^2 dx - \frac{1}{2}\rho g(l - x)\left(\frac{\partial y}{\partial x}\right)^2\right] dx dt, \tag{21}$$

provided λ be now replaced in V by l which differs but slightly from it.

Hamilton's Principle states that (21) must be a maximum or minimum and the integrand is of precisely the form (19) except for a change of notation. Hence

$$-\frac{d}{dx}\left[-\rho g\,(l - x)\frac{\partial y}{\partial x}\right] - \frac{d}{dt}\left(\rho\frac{\partial y}{\partial t}\right) = 0 \quad \text{or} \quad \frac{1}{g}\frac{\partial^2 y}{\partial t^2} = (l - x)\frac{\partial^2 y}{\partial x^2} - \frac{\partial y}{\partial x}.$$

The change of variable $l - x = u^2$, which brings the origin to the end of the chain and reverses the direction of the axis, gives the differential equation

$$\frac{\partial^2 y}{\partial u^2} + \frac{1}{u}\frac{\partial y}{\partial u} = \frac{4}{g}\frac{\partial^2 y}{\partial t^2} \quad \text{or} \quad \frac{d^2 P}{du^2} + \frac{1}{u}\frac{dP}{du} + \frac{4n^2}{g}P = 0 \quad \text{if} \quad y = P(u)\cos nt.$$

As the equation is a partial differential equation the usual device of writing the dependent variable as the product of two functions and trying for a special type of solution has been used (§ 194). The equation in P is a Bessel equation (§ 107) of which one solution $P(u) = AJ_0(2ng^{-\frac{1}{2}}u)$ is finite at the origin $u = 0$, while the other is infinite and must be discarded as not representing possible motions. Thus

$$y(x, t) = AJ_0(2ng^{-\frac{1}{2}}u)\cos nt, \quad \text{with} \quad y(l, t) = AJ_0(2ng^{-\frac{1}{2}}l^{\frac{1}{2}}) = 0$$

as the condition that the chain shall be tied at the original origin, is a possible mode of motion for the chain and consists of whipping back and forth in the periodic time $2\pi/n$. The condition $J_0(2ng^{-\frac{1}{2}}l^{\frac{1}{2}}) = 0$ limits n to one of an infinite set of values obtained from the roots of J_0.

Let there be found the equations for the motion of a medium in which

$$T = \frac{1}{2}A\iiint\left[\left(\frac{\partial\xi}{\partial t}\right)^2 + \left(\frac{\partial\eta}{\partial t}\right)^2 + \left(\frac{\partial\zeta}{\partial t}\right)^2\right] dxdydz,$$

$$V = \frac{1}{2}B\iiint\left(f^2 + g^2 + h^2\right) dxdydz$$

are the kinetic and potential energies, where A and B are constants and

$$4\pi f = \frac{\partial\zeta}{\partial y} - \frac{\partial\eta}{\partial z}, \qquad 4\pi g = \frac{\partial\xi}{\partial z} - \frac{\partial\zeta}{\partial x}, \qquad 4\pi h = \frac{\partial\eta}{\partial x} - \frac{\partial\xi}{\partial y}.$$

are relations connecting f, g, h with the displacements ξ, η, ζ along the axes of x, y, z. Then

$$\iiiint \delta[\tfrac{1}{2}A(\dot\xi^2 + \dot\eta^2 + \dot\zeta^2) - \tfrac{1}{2}B(f^2 + g^2 + h^2)]\,dxdydzdt = 0 \qquad (22)$$

is the expression of Hamilton's Principle. These integrals are more general than (19), for there are three dependent variables ξ, η, ζ and four independent variables x, y, z, t of which they are functions. It is therefore necessary to apply the method of variations directly.

After taking the variations an integration by parts will be applied to the variation of each derivative and the integrated terms will be discarded.

$$\iiiint \delta\,\tfrac{1}{2}A(\dot\xi^2 + \dot\eta^2 + \dot\zeta^2)\,dxdydzdt = \iiiint A(\dot\xi\delta\dot\xi + \dot\eta\delta\dot\eta + \dot\zeta\delta\dot\zeta)\,dxdydzdt$$

$$= -\iiiint A(\ddot\xi\delta\xi + \ddot\eta\delta\eta + \ddot\zeta\delta\zeta)\,dxdydzdt.$$

$$\iiiint \delta\,\tfrac{1}{2}B(f^2 + g^2 + h^2)\,dxdydzdt = \iiiint B(f\delta f + g\delta g + h\delta h)\,dxdydzdt$$

$$= \iiiint \frac{B}{4\pi}\left[f\left(\frac{\partial\delta\zeta}{\partial y} - \frac{\partial\delta\eta}{\partial z}\right) + g\left(\frac{\partial\delta\xi}{\partial z} - \frac{\partial\delta\zeta}{\partial x}\right) + h\left(\frac{\partial\delta\eta}{\partial x} - \frac{\partial\delta\xi}{\partial y}\right)\right]dxdydzdt$$

$$= -\iiiint \frac{B}{4\pi}\left[\left(\frac{\partial g}{\partial z} - \frac{\partial h}{\partial y}\right)\delta\xi + \left(\frac{\partial h}{\partial x} - \frac{\partial f}{\partial z}\right)\delta\eta + \left(\frac{\partial f}{\partial y} - \frac{\partial g}{\partial x}\right)\delta\zeta\right]dxdydzdt.$$

After substitution in (22) the coefficients of $\delta\xi, \delta\eta, \delta\zeta$ may be severally equated to zero because $\delta\xi, \delta\eta, \delta\zeta$ are each arbitrary. Hence the equations

$$4\pi A\frac{\partial^2\xi}{\partial t^2} = -B\left(\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}\right),\quad 4\pi A\frac{\partial^2\eta}{\partial t^2} = -B\left(\frac{\partial f}{\partial z} - \frac{\partial h}{\partial x}\right),\quad 4\pi A\frac{\partial^2\zeta}{\partial t^2} = -B\left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right).$$

With the proper determination of A and B and the proper interpretation of $\xi, \eta, \zeta, f, g, h$, these are the equations of electromagnetism for the free ether.

EXERCISES

1. Show that the straight line is the shortest line in space and that the shortest distance between two curves or surfaces will be normal to both.

2. If at each point of a curve on a surface a geodesic be erected perpendicular to the curve, the locus of its extremity is perpendicular to the geodesic.

3. With any two points of a surface as foci construct a geodesic ellipse by taking the distances $FP + F'P = 2a$ along the geodesics. Show that the tangent to the ellipse is equally inclined to the two geodesic focal radii.

4. Extend Ex. 2, p. 408, to space. If $\int_O^P F(x, y, z)\,ds = \text{const.}$, show that the locus of P is a surface normal to the radii, provided the radii be curves which make the integral a maximum or minimum.

5. Obtain the polar equations for the motion of a particle in a plane.

6. Find the polar equations for the motion of a particle in space.

7. A particle glides down a helicoid ($z = k\phi$ in cylindrical coördinates). Find the equations of motion in (r, ϕ), (r, z), or (z, ϕ), and carry the integration as far as possible toward expressing the position as a function of the time.

8. If $z = ax^2 + by^2 + \cdots$, with $a > 0$, $b > 0$, is the Maclaurin expansion of a surface tangent to the plane $z = 0$ at $(0, 0)$, find and solve the equations for the motion of a particle gliding about on the surface and remaining near the origin.

9. Show that $r(1 + q^2) + t(1 + p^2) - 2pqs = 0$ is the partial differential equation of a minimum surface; test the helicoid.

10. If ρ and S are the density and tension in a uniform piano wire, show that the approximate expressions for the kinetic and potential energies are

$$T = \frac{1}{2}\int_0^l \rho\left(\frac{\partial y}{\partial t}\right)^2 dx, \qquad V = \frac{1}{2}\int_0^l S\left(\frac{\partial y}{\partial x}\right)^2 dx.$$

Obtain the differential equation of the motion and try for solutions $y = P(x)\cos nt$.

11. If ξ, η, ζ are the displacements in a uniform elastic medium, and

$$a = \frac{\partial \xi}{\partial x}, \quad b = \frac{\partial \eta}{\partial y}, \quad c = \frac{\partial \zeta}{\partial z}, \quad f = \left(\frac{\partial \zeta}{\partial y} + \frac{\partial \eta}{\partial z}\right), \quad g = \left(\frac{\partial \xi}{\partial z} + \frac{\partial \zeta}{\partial x}\right), \quad h = \left(\frac{\partial \eta}{\partial x} + \frac{\partial \xi}{\partial y}\right)$$

are six combinations of the nine possible first partial derivatives, it is assumed that $V = \iiint F dx dy dz$, where F is a homogeneous quadratic function of a, b, c, f, g, h, with constant coefficients. Establish the equations of the motion of the medium.

$$\rho\frac{\partial^2 \xi}{\partial t^2} = \frac{\partial^2 F}{\partial x \partial a} + \frac{\partial^2 F}{\partial y \partial h} + \frac{\partial^2 F}{\partial z \partial g}, \qquad \rho\frac{\partial^2 \eta}{\partial t^2} = \frac{\partial^2 F}{\partial x \partial h} + \frac{\partial^2 F}{\partial y \partial b} + \frac{\partial^2 F}{\partial z \partial f},$$

$$\rho\frac{\partial^2 \zeta}{\partial t^2} = \frac{\partial^2 F}{\partial x \partial g} + \frac{\partial^2 F}{\partial y \partial f} + \frac{\partial^2 F}{\partial z \partial c}.$$

12. Establish the conditions (11) by the method of the text in § 155.

13. By the method of § 159 and footnote establish the conditions at the end points for a minimum of $\int F(x, y, y')\,dx$ in terms of F instead of Φ.

14. Prove Stokes's Formula $I = \int_{\bigcirc} \mathbf{F}\cdot d\mathbf{r} = \iint \nabla\times\mathbf{F}\cdot d\mathbf{S}$ of p. 345 by the calculus of variations along the following lines: First compute the variation of I on passing from one closed curve to a neighboring (larger) one.

$$\delta I = \delta\int_{\bigcirc} \mathbf{F}\cdot d\mathbf{r} = \int_{\bigcirc}(\delta\mathbf{F}\cdot d\mathbf{r} - d\mathbf{F}\cdot\delta\mathbf{r}) + \int_{\bigcirc} d(\mathbf{F}\cdot\delta\mathbf{r}) = \int_{\bigcirc}(\nabla\times\mathbf{F})\cdot(\delta\mathbf{r}\times d\mathbf{r}),$$

where the integral of $d(\mathbf{F}\cdot\delta\mathbf{r})$ vanishes. Second interpret the last expression as the integral of $\nabla\times\mathbf{F}\cdot d\mathbf{S}$ over the ring formed by one position of the closed curve and a neighboring position. Finally sum up the variations δI which thus arise on passing through a succession of closed curves expanding from a point to final coincidence with the given closed curve.

15. In case the integrand contains y'' show by successive integrations by parts that

$$\delta\int_{x_0}^{x_1} F(x, y, y', y'')\,dx = \left[Y'\omega + Y''\omega' - \frac{dY''}{dx}\omega\right]_0^1 + \int_{x_0}^{x_1}\left(Y - \frac{dY'}{dx} + \frac{d^2 Y''}{dx^2}\right)\omega\,dx,$$

where $\qquad Y = \dfrac{\partial F}{\partial y}, \qquad Y' = \dfrac{\partial F}{\partial y'}, \qquad Y'' = \dfrac{\partial F}{\partial y''}, \qquad \omega = \delta y.$

PART IV. THEORY OF FUNCTIONS

CHAPTER XVI

INFINITE SERIES

162. Convergence or divergence of series.* Let a series

$$\sum_0^\infty u = u_0 + u_1 + u_2 + \cdots + u_{n-1} + u_n + \cdots, \tag{1}$$

the terms of which are constant but infinite in number, be given. Let the sum of the first n terms of the series be written

$$S_n = u_0 + u_1 + u_2 + \cdots + u_{n-1} = \sum_0^{n-1} u. \tag{2}$$

Then $$S_1,\ S_2,\ S_3,\ \cdots,\ S_n,\ S_{n+1},\ \cdots$$

form a definite suite of numbers which *may approach a definite limit* $\lim S_n = S$ when n becomes infinite. In this case the series is said to *converge to the value* S, and S, which is the limit of the sum of the first n terms, is called the *sum* of the series. Or S_n *may not approach a limit* when n becomes infinite, either because the values of S_n become infinite or because, though remaining finite, they oscillate about and fail to settle down and remain in the vicinity of a definite value. In these cases the series is said to *diverge.*

The necessary and sufficient condition that a series converge is that a value of n may be found so large that the numerical value of $S_{n+p} - S_n$ shall be less than any assigned value for every value of p. (See § 21, Theorem 3, and compare p. 356.) A sufficient condition that a series diverge is that the terms u_n do not approach the limit 0 when n becomes infinite. For if there are always terms numerically as great as some number r no matter how far one goes out in the series, there must always be successive values of S_n which differ by as much as r no matter how large n, and hence the values of S_n cannot possibly settle down and remain in the vicinity of some definite limiting value S.

* It will be useful to read over Chap. II, §§ 18–22, and Exercises. It is also advisable to compare many of the results for infinite series with the corresponding results for infinite integrals (Chap. XIII).

A series in which the terms are alternately positive and negative is called an *alternating series. An alternating series in which the terms approach* 0 *as a limit when n becomes infinite, each term being less than its predecessor, will converge and the difference between the sum S of the series and the sum S_n of the first n terms is less than the next term u_n.* This follows (p. 39, Ex. 3) from the fact that $|S_{n+p} - S_n| < u_n$ and $u_n \doteq 0$.

For example, consider the alternating series

$$1 - x^2 + 2x^4 - 3x^6 + \cdots + (-1)^n nx^{2n} + \cdots.$$

If $|x| \geqq 1$, the individual terms in the series do not approach 0 as n becomes infinite and the series diverges. If $|x| < 1$, the individual terms do approach 0; for

$$\lim_{n\doteq\infty} nx^{2n} = \lim_{n\doteq\infty} \frac{n}{x^{-2n}} = \lim_{n\doteq\infty} \frac{1}{-2x^{-2n}\log x} = 0.$$

And for sufficiently large* values of n the successive terms decrease in magnitude since

$$nx^{2n} < (n-1)x^{2n-2} \quad \text{gives} \quad \frac{n-1}{n} > x^2 \quad \text{or} \quad n > \frac{1}{1-x^2}.$$

Hence the series is seen to converge for any value of x numerically less than unity and to diverge for all other values.

THE COMPARISON TEST. *If the terms of a series are all positive (or all negative) and each term is numerically less than the corresponding term of a series of positive terms which is known to converge, the series converges and the difference $S - S_n$ is less than the corresponding difference for the series known to converge.* (Cf. p. 355.) Let

$$u_0 + u_1 + u_2 + \cdots + u_{n-1} + u_n + \cdots$$

and
$$u'_0 + u'_1 + u'_2 + \cdots + u'_{n-1} + u'_n + \cdots$$

be respectively the given series and the series known to converge. Since the terms of the first are less than those of the second,

$$S_{n+p} - S_n = u_n + \cdots + u_{n+p-1} < u'_n + \cdots + u'_{n+p-1} = S'_{n+p} - S'_n.$$

Now as the second quantity $S'_{n+p} - S'_n$ can be made as small as desired, so can the first quantity $S_{n+p} - S_n$, which is less; and the series must converge. The remainders

$$R_n = S - S_n = u_n + u_{n+1} + \cdots = \sum_n^\infty u,$$

$$R'_n = S' - S'_n = u'_n + u'_{n+1} + \cdots = \sum_n^\infty u'$$

* It should be remarked that the behavior of a series near its beginning is of no consequence in regard to its convergence or divergence; the first N terms may be added and considered as a finite sum S_N and the series may be written as $S_N + u_N + u_{N+1} + \cdots$; it is the properties of $u_N + u_{N+1} + \cdots$ which are important, that is, the ultimate behavior of the series.

clearly satisfy the stated relation $R_n < R'_n$. The series which is most frequently used for comparison with a given series is the geometric,

$$a + ar + ar^2 + ar^3 + \cdots, \qquad R_n = \frac{ar^n}{1-r}, \qquad 0 < r < 1, \tag{3}$$

which is known to converge for all values of r less than 1.

For example, consider the series

$$1 + 1 + \frac{1}{2} + \frac{1}{2\cdot 3} + \frac{1}{2\cdot 3\cdot 4} + \cdots + \frac{1}{n\,!} + \cdots$$

and
$$1 + \frac{1}{2} + \frac{1}{2\cdot 2} + \frac{1}{2\cdot 2\cdot 2} + \cdots + \frac{1}{2^{n-1}} + \cdots.$$

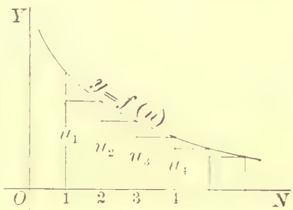
Here, after the first two terms of the first and the first term of the second, each term of the second is greater than the corresponding term of the first. Hence the first series converges and the remainder after the term $1/n\,!$ is less than

$$R_n < \frac{1}{2^n} + \frac{1}{2^{n+1}} + \cdots = \frac{1}{2^n}\,\frac{1}{1-\frac{1}{2}} = \frac{1}{2^{n-1}}.$$

A better estimate of the remainder after the term $1/n\,!$ may be had by comparing

$$R_n = \frac{1}{(n+1)\,!} + \frac{1}{(n+2)\,!} + \cdots \quad \text{with} \quad \frac{1}{(n+1)\,!} + \frac{1}{(n+1)\,!\,(n+1)} + \cdots = \frac{1}{n\,!\,n}.$$

163. As the convergence and divergence of a series are of vital importance, it is advisable to have a number of tests for the convergence or divergence of a given series. The test by comparison with a series known to converge requires that at least a few types of convergent series be known. For the establishment of such types and for the test of many series, the terms of which are positive, *Cauchy's integral test* is useful. Suppose that the terms of the series are decreasing and that a function $f(n)$ which decreases can be found such that $u_n = f(n)$. Now if the terms u_n be plotted at unit intervals along the n-axis, the value of the terms may be interpreted as the area of certain rectangles. The curve $y = f(n)$ lies above the rectangles and the area under the curve is



$$\int_1^n f(n)\,dn > u_2 + u_3 + \cdots + u_n. \tag{4}$$

Hence if the integral converges (which in practice means that if

$$\int f(n)\,dn = F(n), \quad \text{then} \quad \int_1^{x} f(n) = F(x) - F(1) \text{ is finite}\Big),$$

it follows that the series must converge. For instance, if

$$\frac{1}{1^p}+\frac{1}{2^p}+\frac{1}{3^p}+\cdots+\frac{1}{n^p}+\cdots \tag{5}$$

be given, then $u_n=f(n)=1/n^p$, and from the integral test

$$\frac{1}{2^p}+\frac{1}{3^p}+\cdots<\int_1^\infty\frac{dn}{n^p}=\frac{-1}{(p-1)\,n^{p-1}}\bigg]_1^\infty=\frac{1}{p-1}$$

provided $p>1$. Hence the series converges if $p>1$. This series is also very useful for comparison with others; it diverges if $p \leqq 1$ (see Ex. 8).

The Ratio Test. *If the ratio of two successive terms in a series of positive terms approaches a limit which is less than* 1, *the series converges:* if the ratio approaches a limit which is greater than one or if the ratio becomes infinite, the series diverges. That is

if $$\lim_{n=\infty}\frac{u_{n+1}}{u_n}=\gamma<1, \text{ the series converges,}$$

if $$\lim_{n=\infty}\frac{u_{n+1}}{u_n}=\gamma'>1, \text{ the series diverges.}$$

For in the first case, as the ratio approaches a limit less than 1, it must be possible to go so far in the series that the ratio shall be as near to $\gamma<1$ as desired, and hence shall be less than r if r is an assigned number between γ and 1. Then

$$u_{n+1}<ru_n,\qquad u_{n+2}<ru_{n+1}<r^2u_n,\ \cdots$$

and $$u_n+u_{n+1}+u_{n+2}+\cdots<u_n(1+r+r^2+\cdots)=u_n\frac{1}{1-r}.$$

The proof of the divergence when u_{n+1}/u_n becomes infinite or approaches a limit greater than 1 consists in noting that the individual terms cannot approach 0. Note that if the limit of the ratio is 1, *no information* relative to the convergence or divergence is furnished by this test.

If the series of numerical or absolute values

$$|u_0|+|u_1|+|u_2|+\cdots+|u_n|+\cdots$$

of the terms of a series which contains positive and negative terms converges, the series converges and is said to *converge absolutely.* For consider the two sums

$$S_{n+p}-S_n=u_n+\cdots+u_{n+p-1}\quad\text{and}\quad|u_n|+\cdots+|u_{n+p-1}|.$$

The first is surely not numerically greater than the second; as the second can be made as small as desired, so can the first. It follows therefore that the given series must converge. The converse proposition

that if a series of positive and negative terms converges, then the series of absolute values converges, is not true.

As an example on convergence consider the binomial series

$$1 + mx + \frac{m(m-1)}{1 \cdot 2}x^2 + \frac{m(m-1)(m-2)}{1 \cdot 2 \cdot 3}x^3 + \cdots + \frac{m(m-1)\cdots(m-n+1)}{1 \cdot 2 \cdots n}x^n + \cdots,$$

where $$\frac{|u_{n+1}|}{|u_n|} = \frac{|m-n|}{n+1}|x|, \qquad \lim_{n=\infty}\frac{|u_{n+1}|}{|u_n|} = |x|.$$

It is therefore seen that the limit of the quotient of two successive terms in the series of absolute values is $|x|$. This is less than 1 for values of x numerically less than 1, and hence for such values the series converges and converges absolutely. (That the series converges for *positive* values of x less than 1 follows from the fact that for values of n greater than $m + 1$ the series alternates and the terms approach 0; the proof above holds equally for negative values.) For values of x numerically greater than 1 the series does not converge absolutely. As a matter of fact when $|x| > 1$, the series does not converge at all; for as the ratio of successive terms approaches a limit greater than unity, the individual terms cannot approach 0. For the values $x = \pm 1$ the test fails to give information. The conclusions are therefore that for values of $|x| < 1$ the binomial series converges absolutely, for values of $|x| > 1$ it diverges, and for $|x| = 1$ the question remains doubtful.

A word about series with *complex terms*. Let

$$\begin{aligned} &u_0 + u_1 + u_2 + \cdots + u_{n-1} + u_n + \cdots \\ &= u_0' + u_1' + u_2' + \cdots + u_{n-1}' + u_n' + \cdots \\ &\quad + i(u_0'' + u_1'' + u_2'' + \cdots + u_{n-1}'' + u_n'' + \cdots) \end{aligned}$$

be a series of complex terms. The sum to n terms is $S_n = S_n' + iS_n''$. The series is said to converge if S_n approaches a limit when n becomes infinite. If the complex number S_n is to approach a limit, both its real part S_n' and the coefficient S_n'' of its imaginary part must approach limits, and hence the series of real parts and the series of imaginary parts must converge. It will then be possible to take n so large that for any value of p the simultaneous inequalities

$$|S_{n+p}' - S_n'| < \tfrac{1}{2}\epsilon \quad \text{and} \quad |S_{n+p}'' - S_n''| < \tfrac{1}{2}\epsilon,$$

where ϵ is any assigned number, hold. Therefore

$$|S_{n+p} - S_n| \leqq |S_{n+p}' - S_n'| + |iS_{n+p}'' - iS_n''| < \epsilon.$$

Hence if the series converges, the same condition holds as for a series of real terms. Now conversely the condition

$$|S_{n+p} - S_n| < \epsilon \quad \text{implies} \quad |S_{n+p}' - S_n'| < \epsilon, \qquad |S_{n+p}'' - S_n''| < \epsilon.$$

Hence if the condition holds, the two real series converge and the complex series will then converge.

164. As Cauchy's integral test is not easy to apply except in simple cases and the ratio test fails when the limit of the ratio is 1, other sharper tests for convergence or divergence are sometimes needed, as in the case of the binomial series when $x = \pm 1$. Let there be given two series of positive terms

$$u_0 + u_1 + \cdots + u_n + \cdots \quad \text{and} \quad v_0 + v_1 + \cdots + v_n + \cdots$$

of which the first is to be tested and the second is known to converge (or diverge). *If the ratio of two successive terms u_{n+1}/u_n ultimately becomes and remains less* (*or greater*) *than the ratio v_{n+1}/v_n, the first series is also convergent* (*or divergent*). For if

$$\frac{u_{n+1}}{u_n} < \frac{v_{n+1}}{v_n}, \qquad \frac{u_{n+2}}{u_{n+1}} < \frac{v_{n+2}}{v_{n+1}}, \qquad \cdots, \quad \text{then} \quad \frac{u_n}{v_n} > \frac{u_{n+1}}{v_{n+1}} > \frac{u_{n+2}}{v_{n+2}} > \cdots.$$

Hence if $\quad u_n = \rho v_n, \qquad$ then $\quad u_{n+1} < \rho v_{n+1}, \qquad u_{n+2} < \rho v_{n+2}, \qquad \cdots,$

and
$$u_n + u_{n+1} + u_{n+2} + \cdots < \rho\,(v_n + v_{n+1} + v_{n+2} + \cdots).$$

As the v-series is known to converge, the ρv-series serves as a comparison series for the u-series which must then converge. If $u_{n+1}/u_n > v_{n+1}/v_n$ and the v-series diverges, similar reasoning would show that the u-series diverges.

This theorem serves to establish the useful *test due to Raabe*, which is

$$\text{if } \lim_{n \doteq \infty} n\left(\frac{u_n}{u_{n+1}} - 1\right) > 1, \ S_n \text{ converges}; \qquad \text{if } \lim_{n \doteq \infty} n\left(\frac{u_n}{u_{n+1}} - 1\right) < 1, \ S_n \text{ diverges}.$$

Again, if the limit is 1, no information is given. This test need never be tried except when the ratio test gives a limit 1 and fails. The proof is simple. For

$$\int^{\infty} \frac{dn}{n(\log n)^{1+\alpha}} = -\frac{1}{\alpha}\frac{1}{(\log n)^{\alpha}}\Bigg]^{\infty} \text{ is finite}$$

and
$$\int^{\infty} \frac{dn}{n \log n} = \log\log n\Bigg]^{\infty} \text{ is infinite,}$$

hence $\displaystyle \frac{1}{2(\log 2)^{1+\alpha}} + \cdots + \frac{1}{n(\log n)^{1+\alpha}} + \cdots \quad$ and $\quad \displaystyle \frac{1}{2(\log 2)} + \cdots + \frac{1}{n(\log n)} + \cdots$

are respectively convergent and divergent by Cauchy's integral test. Let these be taken as the v-series with which to compare the u-series. Then

$$\frac{v_n}{v_{n+1}} = \frac{n+1}{n}\left(\frac{\log(n+1)}{\log n}\right)^{1+\alpha} = \left(1 + \frac{1}{n}\right)\left(\frac{\log(1+n)}{\log n}\right)^{1+\alpha}$$

and
$$\frac{v_n}{v_{n+1}} = \left(1 + \frac{1}{n}\right)\frac{\log(1+n)}{\log n}$$

in the two respective cases. Next consider Raabe's expression. If first

$$\lim_{n \doteq \infty} n\left(\frac{u_n}{u_{n+1}} - 1\right) > 1, \quad \text{then ultimately} \quad n\left(\frac{u_n}{u_{n+1}} - 1\right) > \gamma > 1 \quad \text{and} \quad \frac{u_n}{u_{n+1}} > 1 + \frac{\gamma}{n}.$$

Now $\quad \displaystyle \lim_{n \doteq \infty}\left(\frac{\log(1+n)}{\log n}\right)^{1+\alpha} = 1 \quad$ and ultimately $\quad \displaystyle \left(\frac{\log(1+n)}{\log n}\right)^{1+\alpha} < 1 + \epsilon,$

where ϵ is arbitrarily small. Hence ultimately if $\gamma > 1$,

$$\left(1+\frac{1}{n}\right)\left(\frac{\log(1+n)}{\log n}\right)^{1+\alpha} < 1 + \frac{1+\epsilon}{n} + \frac{\epsilon}{n^2} < 1 + \frac{\gamma}{n},$$

or $$v_n/v_{n+1} < u_n/u_{n+1} \quad \text{or} \quad u_{n+1}/u_n < v_{n+1}/v_n,$$

and the u-series converges. In like manner, secondly, if

$$\lim_{n=\infty} n\left(\frac{u_n}{u_{n+1}} - 1\right) < 1, \quad \text{then ultimately} \quad \frac{u_n}{u_{n+1}} < 1 + \frac{\gamma}{n}, \qquad \gamma < 1;$$

and $$1 + \frac{\gamma}{n} < \left(1+\frac{1}{n}\right)\frac{\log(1+n)}{\log n} \quad \text{or} \quad \frac{u_n}{u_{n+1}} < \frac{v_n}{v_{n+1}} \quad \text{or} \quad \frac{u_{n+1}}{u_n} > \frac{v_{n+1}}{v_n}.$$

Hence as the v-series now diverges, the u-series must diverge.

Suppose this test applied to the binomial series for $x = -1$. Then

$$\frac{u_n}{u_{n+1}} = \frac{n+1}{n-m}, \qquad \lim_{n=\infty} n\left(\frac{n+1}{n-m} - 1\right) = \lim_{n=\infty} \frac{m+1}{1-\frac{m}{n}} = m+1.$$

It follows that the series will converge if $m > 0$, but diverge if $m < 0$. If $x = +1$, the binomial series becomes alternating for $n > m + 1$. If the series of absolute values be considered, the ratio of successive terms $|u_n/u_{n+1}|$ is still $(n+1)/(n-m)$ and the binomial series converges absolutely if $m > 0$; but when $m < 0$ the series of absolute values diverges and it remains an open question whether the alternating series diverges or converges. Consider therefore the alternating series

$$1 + m + \frac{m(m-1)}{1\cdot 2} + \frac{m(m-1)(m-2)}{1\cdot 2\cdot 3} + \cdots + \frac{m(m-1)\cdots(m-n+1)}{1\cdot 2\cdots n} + \cdots, \quad m < 0.$$

This will converge if the limit of u_n is 0, but otherwise it will diverge. Now if $m \leqq -1$, the successive terms are multiplied by a factor $|m - n + 1|/n \geqq 1$ and they cannot approach 0. When $-1 < m < 0$, let $1 + m = \theta$, a fraction. Then the nth term in the series is

$$|u_n| = (1-\theta)\left(1-\frac{\theta}{2}\right)\cdots\left(1-\frac{\theta}{n}\right)$$

and $$-\log|u_n| = -\log(1-\theta) - \log\left(1-\frac{\theta}{2}\right) - \cdots - \log\left(1-\frac{\theta}{n}\right).$$

Each successive factor diminishes the term but diminishes it by so little that it may not approach 0. The logarithm of the term is a series. Now apply Cauchy's test.

$$\int^{\infty} -\log\left(1-\frac{\theta}{n}\right)dn = \left[-n\log\left(1-\frac{\theta}{n}\right) + \theta\log(n-\theta)\right]^{\infty} = \infty.$$

The series of logarithms therefore diverges and $\lim|u_n| = e^{-\infty} = 0$. Hence the terms approach 0 as a limit. The final results are therefore that when $x = -1$ the binomial series converges if $m > 0$ but diverges if $m < 0$; and when $x = +1$ it converges (absolutely) if $m > 0$, diverges if $m < -1$, and converges (not absolutely) if $-1 < m < 0$.

EXERCISES

1. State the number of terms which must be taken in these alternating series to obtain the sum accurate to three decimals. If the number is not greater than 8, compute the value of the series to three decimals, carrying four figures in the work:

$(\alpha)\ \frac{1}{3}-\frac{1}{2\cdot 3^2}+\frac{1}{3\cdot 3^3}-\frac{1}{4\cdot 3^4}+\cdots,$ $\qquad(\beta)\ \frac{1}{2}-\frac{1}{2\cdot 2^2}+\frac{1}{3\cdot 2^3}-\frac{1}{4\cdot 2^4}+\cdots,$

$(\gamma)\ 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots,$ $\qquad(\delta)\ \frac{1}{\log 2}-\frac{1}{\log 3}+\frac{1}{\log 4}-\cdots,$

$(\epsilon)\ 1-\frac{1}{3^2}+\frac{1}{5^2}-\frac{1}{7^2}+\cdots,$ $\qquad(\zeta)\ e^{-1}-2e^{-2}+3e^{-3}-4e^{-4}+\cdots.$

2. Find the values of x for which these alternating series converge or diverge:

$(\alpha)\ 1-x^2+\frac{1}{2}x^4-\frac{1}{3}x^6+\cdots,$ $\qquad(\beta)\ 1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots,$

$(\gamma)\ x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots,$ $\qquad(\delta)\ x-\frac{x^3}{3}+\frac{x^5}{5}-\frac{x^7}{7}+\cdots,$

$(\epsilon)\ 1-\frac{x^2}{1^p}+\frac{x^4}{2^p}-\frac{x^6}{3^p}+\cdots,$ $\qquad(\zeta)\ 2x-\frac{2^3x^3}{3}+\frac{2^5x^5}{5}-\frac{2^7x^7}{7}+\cdots,$

$(\eta)\ \frac{1}{x}-\frac{1}{x+1}+\frac{1}{x+2}-\frac{1}{x+3}+\cdots,$ $\qquad(\theta)\ \frac{1}{x}-\frac{2}{x+1}+\frac{2^2}{x+2}-\frac{2^3}{x+3}+\cdots.$

3. Show that these series converge and estimate the error after n terms:

$(\alpha)\ 1+\frac{1}{2^2}+\frac{1}{3^3}+\frac{1}{4^4}+\cdots,$ $\qquad(\beta)\ \frac{1}{3}+\frac{1\cdot 2}{3\cdot 5}+\frac{1\cdot 2\cdot 3}{3\cdot 5\cdot 7}+\cdots,$

$(\gamma)\ \frac{1}{2}+\frac{1}{2\cdot 2^2}+\frac{1}{3\cdot 2^3}+\frac{1}{4\cdot 2^4}+\cdots,$ $\qquad(\delta)\ \left(\frac{1}{3}\right)^2+\left(\frac{1\cdot 2}{3\cdot 5}\right)^2+\left(\frac{1\cdot 2\cdot 3}{3\cdot 5\cdot 7}\right)^2+\cdots.$

From the estimate of error state how many terms are required to compute the series accurate to two decimals and make the computation, carrying three figures. Test for convergence or divergence:

$(\epsilon)\ \sin 1+\sin\frac{1}{2}+\sin\frac{1}{3}+\cdots,$ $\qquad(\zeta)\ \sin^2 1+\sin^2\frac{1}{2}+\sin^2\frac{1}{3}+\cdots,$

$(\eta)\ \tan^{-1}1+\tan^{-1}\frac{1}{2}+\tan^{-1}\frac{1}{3}+\cdots,$ $\qquad(\theta)\ \tan 1+\frac{1}{\sqrt{2}}\tan\frac{1}{2}+\frac{1}{\sqrt{3}}\tan\frac{1}{3}+\cdots,$

$(\iota)\ \frac{1}{1+1}+\frac{1}{2+\sqrt{2}}+\frac{1}{3+\sqrt{3}}+\cdots,$ $\qquad(\kappa)\ \frac{1}{2^2-1^2}+\frac{1}{3^2-2^2}+\frac{1}{4^2-3^2}+\cdots,$

$(\lambda)\ \frac{1}{x}+\frac{2}{x^2}+\frac{2\cdot 3}{x^3}+\frac{2\cdot 3\cdot 4}{x^4}+\cdots,$ $\qquad(\mu)\ \frac{1}{x}+\frac{\sqrt{2}}{x^2}+\frac{\sqrt[3]{3}}{x^3}+\frac{\sqrt[4]{4}}{x^4}+\cdots.$

4. Apply Cauchy's integral to determine the convergence or divergence:

$(\alpha)\ 1+\frac{\log 2}{2^p}+\frac{\log 3}{3^p}+\frac{\log 4}{4^p}+\cdots,$ $\qquad(\beta)\ 1+\frac{1}{2(\log 2)^p}+\frac{1}{3(\log 3)^p}+\frac{1}{4(\log 4)^p}+\cdots.$

$(\gamma)\ 1+\sum_{2}^{\infty}\frac{1}{n\log n\log\log n}$, $(\delta)\ 1+\sum_{2}^{\infty}\frac{1}{n\log n(\log\log n)^p}$,

$(\epsilon)\ \cot^{-1}1+\cot^{-1}2+\cdots$. $(\zeta)\ 1+\frac{2}{2^2+1}+\frac{3}{3^2+2}+\frac{4}{4^2+3}+\cdots$.

5. Apply the ratio test to determine convergence or divergence:

$(\alpha)\ \frac{1}{2}+\frac{2}{2^2}+\frac{3}{2^3}+\frac{4}{2^4}+\cdots$, $(\beta)\ \frac{2^2}{2^{10}}+\frac{2^3}{3^{10}}+\frac{2^4}{4^{10}}+\cdots$,

$(\gamma)\ \frac{2\,!}{2^5}+\frac{3\,!}{3^5}+\frac{4\,!}{4^5}+\frac{5\,!}{5^5}+\cdots$, $(\delta)\ \frac{2^2}{2\,!}+\frac{3^3}{3\,!}+\frac{4^4}{4\,!}+\cdots$,

(ϵ) Ex. 3 (α), (β), (γ), (δ); Ex. 4 (α), (ζ), $(\zeta)\ \frac{2^{10}}{10^2}+\frac{3^{10}}{10^3}+\frac{4^{10}}{10^4}+\cdots$,

$(\eta)\ 1+\frac{x^2}{2\,!}+\frac{x^4}{4\,!}+\frac{x^6}{6\,!}+\cdots$, $(\theta)\ 1+\frac{x^2}{2^p}+\frac{x^4}{4^p}+\cdots$,

$(\iota)\ x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots$, $(\kappa)\ \frac{1}{a}+\frac{bx}{a^2}+\frac{b^2x^3}{a^3}+\cdots$.

6. Where the ratio test fails, discuss the above exercises by any method.

7. Prove that if a series of decreasing positive terms converges, $\lim nu_n = 0$.

8. Formulate the Cauchy integral test for divergence and check the statement on page 422. The test has been used in the text and in Ex. 4. *Prove* the test.

9. Show that if the ratio test indicates the divergence of the series of absolute values, the series diverges no matter what the distribution of signs may be.

10. Show that if $\sqrt[n]{u_n}$ approaches a limit less than 1, the series (of positive terms) converges; but if $\sqrt[n]{u_n}$ approaches a limit greater than 1, it diverges.

11. If the terms of a convergent series $u_0+u_1+u_2+\cdots$ of positive terms be multiplied respectively by a set of positive numbers $a_0, a_1, a_2, \cdots$ all of which are less than some number G, the resulting series $a_0u_0+a_1u_1+a_2u_2+\cdots$ converges. State the corresponding theorem for divergent series. What if the given series has terms of opposite signs, but converges absolutely?

12. Show that the series $\frac{\sin x}{1^2}-\frac{\sin 2x}{2^2}+\frac{\sin 3x}{3^2}-\frac{\sin 4x}{4^2}+\cdots$ converges absolutely for any value of x, and that the series $1+x\sin\theta+x^2\sin 2\theta+x^3\sin 3\theta+\cdots$ converges absolutely for any x numerically less than 1, no matter what θ may be.

13. If $a_0, a_1, a_2, \cdots$ are any suite of numbers such that $\sqrt[n]{|a_n|}$ approaches a limit less than or equal to 1, show that the series $a_0+a_1x+a_2x^2+\cdots$ converges absolutely for any value of x numerically less than 1. Apply this to show that the following series converge absolutely when $|x|<1$;

$(\alpha)\ 1+\frac{1}{2}x^2+\frac{1\cdot 3}{2\cdot 4}x^4+\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}x^6+\cdots$, $(\beta)\ 1-2x+3x^2-4x^3+\cdots$,

$(\gamma)\ 1+x+2^px^2+3^px^3+4^px^4+\cdots$, $(\delta)\ 1-x\log 1+x^2\log 4-x^3\log 9+\cdots$.

14. Show that in Ex. 10 it will be sufficient for convergence if $\sqrt[n]{u_n}$ becomes and remains less than $\gamma < 1$ without approaching a limit, and sufficient for divergence if there are an infinity of values for n such that $\sqrt[n]{u_n} > 1$. Note a similar generalization in Ex. 13 and state it.

15. If a power series $a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$ converges for $x = X > 0$, it converges absolutely for any x such that $|x| < X$, and the series

$$a_0x + \tfrac{1}{2}a_1x^2 + \tfrac{1}{3}a_2x^3 + \cdots \quad \text{and} \quad a_1 + 2a_2x + 3a_3x^2 + \cdots,$$

obtained by integrating and differentiating term by term, also converge absolutely for any value of x such that $|x| < X$. The same result, by the same proof, holds if the terms $a_0, a_1X, a_2X^2, \cdots$ remain less than a fixed value G.

16. If the ratio of the successive terms in a series of positive terms be regarded as a function of $1/n$ and may be expanded by Maclaurin's Formula to give

$$\frac{u_n}{u_{n+1}} = \alpha + \beta\frac{1}{n} + \frac{\mu}{2}\left(\frac{1}{n}\right)^2, \qquad \mu \text{ remaining finite as } \frac{1}{n} \doteq 0,$$

the series converges if $\alpha > 1$ or $\alpha = 1, \beta > 1$, but diverges if $\alpha < 1$ or $\alpha = 1, \beta \leqq 1$. This test covers most of the series of positive terms which arise in practice. Apply it to various instances in the text and previous exercises. Why are there series to which this test is inapplicable?

17. If $\rho_0, \rho_1, \rho_2, \cdots$ is a decreasing suite of positive numbers approaching a limit λ and $S_0, S_1, S_2, \cdots$ is any limited suite of numbers, that is, numbers such that $|S_n| \leqq G$, show that the series

$$(\rho_0 - \rho_1)S_0 + (\rho_1 - \rho_2)S_1 + (\rho_2 - \rho_3)S_2 + \cdots \text{ converges absolutely,}$$

and
$$\left|\sum_0^r (\rho_n - \rho_{n+1})S_n\right| \leqq G(\rho_0 - \lambda).$$

18. Apply Ex. 17 to show that, $\rho_0, \rho_1, \rho_2, \cdots$ being a decreasing suite, if

$$u_0 + u_1 + u_2 + \cdots \text{ converges,} \qquad \rho_0u_0 + \rho_1u_1 + \rho_2u_2 + \cdots \text{ will converge also.}$$

N.B. $$\rho_0u_0 + \rho_1u_1 + \cdots + \rho_nu_n = \rho_0S_1 + \rho_1(S_2 - S_1) + \cdots + \rho_n(S_{n+1} - S_n)$$
$$= S_1(\rho_0 - \rho_1) + \cdots + S_n(\rho_{n-1} - \rho_n) + \rho_nS_{n+1}.$$

19. Apply Ex. 18 to prove Ex. 15 after showing that $\rho_0u_0 + \rho_1u_1 + \cdots$ must converge absolutely if $\rho_0 + \rho_1 + \cdots$ converges.

20. If $a_1, a_2, a_3, \cdots, a_n$ are n positive numbers less than 1, show that

$$(1 + a_1)(1 + a_2)\cdots(1 + a_n) > 1 + a_1 + a_2 + \cdots + a_n$$

and
$$(1 - a_1)(1 - a_2)\cdots(1 - a_n) > 1 - a_1 - a_2 - \cdots - a_n$$

by induction or any other method. Then since $1 + a_1 < 1/(1 - a_1)$ show that

$$\frac{1}{1 - (a_1 + a_2 + \cdots + a_n)} > (1 + a_1)(1 + a_2)\cdots(1 + a_n) > 1 + (a_1 + a_2 + \cdots + a_n).$$

$$\frac{1}{1 + (a_1 + a_2 + \cdots + a_n)} > (1 - a_1)(1 - a_2)\cdots(1 - a_n) > 1 - (a_1 + a_2 + \cdots + a_n).$$

if $a_1 + a_2 + \cdots + a_n < 1$. Or if $\prod$ be the symbol for a *product*,

$$\left(1-\sum_1^n a\right)^{-1} > \prod_1^n (1+a) > 1+\sum_1^n a, \qquad \left(1+\sum_1^n a\right)^{-1} > \prod_1^n (1-a) > 1-\sum_1^n a.$$

21. Let $\prod_1^\infty (1+u_1)(1+u_2)\cdots(1+u_n)(1+u_{n+1})\cdots$ be an infinite product and let P_n be the product of the first n factors. Show that $|P_{n+p}-P_n|<\epsilon$ is the necessary and sufficient condition that P_n approach a limit when n becomes infinite. Show that u_n must approach 0 as a limit if P_n approaches a limit.

22. In case P_n approaches a limit different from 0, show that if ϵ be assigned, a value of n can be found so large that for any value of p

$$\left|\frac{P_{n+p}}{P_n}-1\right| = \left|\prod_{n+1}^{n+p}(1+u_i)-1\right|<\epsilon \quad\text{or}\quad \prod_{n+1}^{n+p}(1+u_i)=1+\eta, \qquad |\eta|<\epsilon.$$

Conversely show that if this relation holds, P_n must approach a limit other than 0. The *infinite product* is said to *converge* when P_n approaches a limit other than 0; in all other cases it is said to diverge, including the case where $\lim P_n = 0$.

23. By combining Exs. 20 and 22 show that the necessary and sufficient condition that

$$P_n = (1+a_1)(1+a_2)\cdots(1+a_n) \quad\text{and}\quad Q_n = (1-a_1)(1-a_2)\cdots(1-a_n)$$

converge as n becomes infinite is that the series $a_1 + a_2 + \cdots + a_n + \cdots$ shall converge. Note that P_n is increasing and Q_n decreasing. Show that in case Σa diverges, P_n diverges to ∞ and Q_n to 0 (provided ultimately $a_i < 1$).

24. Define absolute convergence for infinite products and show that if a product converges absolutely it converges in its original form.

25. Test these products for convergence, divergence, or absolute convergence:

(α) $\left(1-\frac{1}{2}\right)\left(1-\frac{1}{4}\right)\left(1-\frac{1}{8}\right)\cdots$, (β) $\left(1+\frac{1}{2^2}\right)\left(1+\frac{1}{3^2}\right)\left(1+\frac{1}{4^2}\right)\cdots$,

(γ) $\prod_1^\infty\left[1-\left(\frac{nx}{n+1}\right)^n\right]$, (δ) $(1+x)(1+x^2)(1+x^4)(1+x^8)\cdots$,

(ε) $\left(1-\frac{1}{\log 2}\right)\left(1-\frac{1}{(\log 4)^2}\right)\left(1-\frac{1}{(\log 8)^3}\right)\cdots$, (ζ) $\prod_1^\infty\left[\left(1-\frac{x}{c+n}\right)e^{\frac{x}{n}}\right]$.

26. Given $\frac{\frac12 u^2}{1+u}$ or $\frac12 u^2 < u - \log(1+u) < \frac12 u^2$ or $\frac{\frac12 u^2}{1+u}$ according as u is a positive or negative fraction (see Ex. 29, p. 11). Prove that if Σu_n^2 converges, then

$$u_{n+1}+u_{n+2}+\cdots+u_{n+p}-\log(1+u_{n+1})(1+u_{n+2})\cdots(1+u_{n+p})$$
$$= (S_{n+p}-S_n)-(\log P_{n+p}-\log P_n)$$

can be made as small as desired by taking n large enough regardless of p. Hence prove that if Σu_n^2 converges, $\prod(1+u_n)$ converges if Σu_n does, but diverges to ∞ if Σu_n diverges to $+\infty$, and diverges to 0 if Σu_n diverges to $-\infty$; whereas if Σu_n^2 diverges while Σu_n converges, the product diverges to 0.

27. Apply Ex. 26 to: $(\alpha)\ \left(1+\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1+\frac{1}{4}\right)\left(1-\frac{1}{5}\right)\cdots,$

$(\beta)\ \left(1-\frac{1}{\sqrt{2}}\right)\left(1+\frac{1}{\sqrt{3}}\right)\left(1-\frac{1}{\sqrt{4}}\right)\cdots,\qquad (\gamma)\ \left(1+\frac{x}{1}\right)\left(1-\frac{x^2}{2}\right)\left(1+\frac{x^3}{3}\right)\left(1-\frac{x^4}{4}\right)\cdots$

28. Suppose the integrand $f(x)$ of an infinite integral oscillates as x becomes infinite. What test might be applicable from the construction of an alternating series?

165. Series of functions. If the terms of a series

$$S(x) = u_0(x) + u_1(x) + \cdots + u_n(x) + \cdots \tag{6}$$

are functions of x, the series defines a function $S(x)$ of x for every value of x for which it converges. If the individual terms of the series are continuous functions of x over some interval $a \leqq x \leqq b$, the sum $S_n(x)$ of n terms will of course be a continuous function over that interval. Suppose that the series converges for all points of the interval. Will it then be true that $S(x)$, the limit of $S_n(x)$, is also a continuous function over the interval? Will it be true that the integral term by term,

$$\int_a^b u_0(x)\,dx + \int_a^b u_1(x)\,dx + \cdots,\quad \text{converges to}\quad \int_a^b S(x)\,dx?$$

Will it be true that the derivative term by term,

$$u_0'(x) + u_1'(x) + \cdots,\quad \text{converges to}\quad S'(x)?$$

There is no *a priori* reason why any of these things should be true; for the proofs which were given in the case of finite sums will not apply to the case of a limit of a sum of an infinite number of terms (cf. § 144).

These questions may readily be thrown into the form of questions concerning the possibility of inverting the order of two limits (see § 44).

For integration: Is $\displaystyle\int_a^b \lim_{n=\infty} S_n(x)\,dx = \lim_{n=\infty}\int_a^b S_n(x)\,dx$?

For differentiation: Is $\displaystyle\frac{d}{dx}\lim_{n=\infty} S_n(x) = \lim_{n=\infty}\frac{d}{dx}S_n(x)$?

For continuity: Is $\displaystyle\lim_{x\doteq x_0}\lim_{n=\infty} S_n(x) = \lim_{n=\infty}\lim_{x\doteq x_0} S_n(x)$?

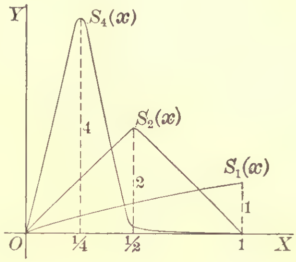
As derivatives and definite integrals are themselves defined as limits, the existence of a double limit is clear. That all three of the questions must be answered in the negative unless some restriction is placed on the way in which $S_n(x)$ converges to $S(x)$ is clear from some examples. Let $0 \leqq x \leqq 1$ and

$$S_n(x) = xn^2e^{-nx},\quad \text{then}\quad \lim_{n=\infty} S_n(x) = 0,\quad \text{or}\quad S(x) = 0.$$

No matter what the value of x, the limit of $S_n(x)$ is 0. The limiting function is therefore continuous in this case; but from the manner in which $S_n(x)$ converges

to $S(x)$ it is apparent that under suitable conditions the limit would not be continuous. The area under the limit $S(x) = 0$ from 0 to 1 is of course 0; but the limit of the area under $S_n(x)$ is

$$\lim_{n=\infty}\int_0^1 xn^2e^{-nx}dx = \lim_{n=\infty}\Big[e^{-nx}(-nx-1)\Big]_0^1 = 1.$$



The derivative of the limit at the point $x = 0$ is of course 0; but the limit,

$$\lim_{n=\infty}\left[\frac{d}{dx}(xn^2e^{-nx})\right]_{x=0}$$

$$= \lim_{n=\infty}\Big[n^2e^{-nx}(1-nx)\Big]_{x=0} = \lim_{n=\infty} n^2 = \infty,$$

of the derivative is infinite. Hence in this case two of the questions have negative answers and one of them a positive answer.

If a suite of functions such as $S_1(x)$, $S_2(x)$, $\cdots$, $S_n(x)$, $\cdots$ converge to a limit $S(x)$ over an interval $a \leqq x \leqq b$, the conception of a limit requires that when ϵ is assigned and x_0 is assumed it must be possible to take n so large that $|R_n(x_0)| = |S(x_0) - S_n(x_0)| < \epsilon$ for this and any larger n. The suite is said to *converge uniformly* toward its limit, if this condition can be satisfied simultaneously for all values of x in the interval, that is, if when ϵ is assigned it is possible to take n so large that $|R_n(x)| < \epsilon$ for every value of x in the interval and for this and any larger n. In the above example the convergence was not uniform; the figure shows that no matter how great n, there are always values of x between 0 and 1 for which $S_n(x)$ departs by a large amount from its limit 0.

The uniform convergence of a continuous function $S_n(x)$ to its limit is sufficient to insure the continuity of the limit $S(x)$. To show that $S(x)$ is continuous it is merely necessary to show that when ϵ is assigned it is possible to find a Δx so small that $|S(x+\Delta x) - S(x)| < \epsilon$. But $|S(x+\Delta x) - S(x)| = |S_n(x+\Delta x) - S_n(x) + R_n(x+\Delta x) - R_n(x)|$; and as by hypothesis R_n converges uniformly to 0, it is possible to take n so large that $|R_n(x+\Delta x)|$ and $|R_n(x)|$ are less than $\frac{1}{3}\epsilon$ irrespective of x. Moreover, as $S_n(x)$ is continuous it is possible to take Δx so small that $|S_n(x+\Delta x) - S_n(x)| < \frac{1}{3}\epsilon$ irrespective of x. Hence $|S(x+\Delta x) - S(x)| < \epsilon$, and the theorem is proved. Although the uniform convergence of S_n to S is a sufficient condition for the continuity of S, it is not a necessary condition, as the above example shows.

The uniform convergence of $S_n(x)$ to its limit insures that

$$\lim_{n=\infty}\int_a^b S_n(x)\,dx = \int_a^b S(x)\,dx.$$

For in the first place $S(x)$ must be continuous and therefore integrable. And in the second place when ϵ is assigned, n may be taken so large that $|R_n(x)| < \epsilon/(b-a)$. Hence

$$\left|\int_a^b S(x)\,dx - \int_a^b S_n(x)\,dx\right| = \left|\int_a^b R_n(x)\,dx\right| < \int_a^b \frac{\epsilon}{b-a}\,dx = \epsilon,$$

and the result is proved. Similarly *if $S_n'(x)$ is continuous and converges uniformly to a limit $T(x)$, then $T(x) = S'(x)$.* For by the above result on integrals,

$$\int_a^x T(x)\,dx = \lim_{n=\infty}\int_a^x S_n'(x)\,dx = \lim_{n=\infty}\Big[S_n(x) - S_n(a)\Big] = S(x) - S(a).$$

Hence $T(x) = S'(x)$. It should be noted that this proves incidentally that if $S_n'(x)$ is continuous and converges uniformly to a limit, then $S(x)$ actually has a derivative, namely $T(x)$.

In order to apply these results to a series, it is necessary to have a *test for the uniformity of the convergence of the series*; that is, for the uniform convergence of $S_n(x)$ to $S(x)$. One such test is ***Weierstrass's M-test***: *The series*

$$u_0(x) + u_1(x) + \cdots + u_n(x) + \cdots \tag{7}$$

will converge uniformly provided a convergent series

$$M_0 + M_1 + \cdots + M_n + \cdots \tag{8}$$

of positive terms may be found such that ultimately $|u_i(x)| \leqq M_i$. The proof is immediate. For

$$|R_n(x)| = |u_n(x) + u_{n+1}(x) + \cdots| \leqq M_n + M_{n+1} + \cdots$$

and as the M-series converges, its remainder can be made as small as desired by taking n sufficiently large. Hence any series of continuous functions defines a continuous function and may be integrated term by term to find the integral of that function provided an M-test series may be found; and the derivative of that function is the derivative of the series term by term if this derivative series admits an M-test.

To apply the work to an example consider whether the series

$$S(x) = \frac{\cos x}{1^2} + \frac{\cos 2x}{2^2} + \frac{\cos 3x}{3^2} + \cdots + \frac{\cos nx}{n^2} + \cdots \tag{7'}$$

defines a continuous function and may be integrated and differentiated term by term as

$$\int_0^x S(x) = \frac{\sin x}{1^3} + \frac{\sin 2x}{2^3} + \frac{\sin 3x}{3^3} + \cdots + \frac{\sin nx}{n^3} + \cdots \tag{7''}$$

and

$$\frac{d}{dx}S(x) = -\frac{\sin x}{1} - \frac{\sin 2x}{2} - \frac{\sin 3x}{3} - \cdots - \frac{\sin nx}{n} - \cdots. \tag{7'''}$$

As $|\cos x| \leqq 1$, the convergent series $1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots$ may be taken as an M-series for $S(x)$. Hence $S(x)$ is a continuous function of x for all real values of x, and the integral of $S(x)$ may be taken as the limit of the integral of $S_n(x)$, that is, as the integral of the series term by term as written. On the other hand, an M-series for (7‴) cannot be found, for the series $1 + \frac{1}{2} + \frac{1}{3} + \cdots$ is not convergent. It therefore appears that $S'(x)$ may not be identical with the term-by-term derivative of $S(x)$; it does not follow that it will not be, — merely that it may not be.

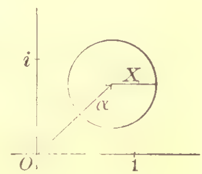
166. Of series with variable terms, the *power series*

$$f(z) = a_0 + a_1(z - \alpha) + a_2(z - \alpha)^2 + \cdots + a_n(z - \alpha)^n + \cdots \tag{9}$$

is perhaps the most important. Here z, α, and the coefficients a_i may be either real or complex numbers. This series may be written more simply by setting $x = z - \alpha$; then

$$f(x + \alpha) = \phi(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots \tag{9'}$$

is a series which surely converges for $x = 0$. It may or may not converge for other values of x, but from Ex. 15 or 19 above it is seen that if the series converges for X, it converges absolutely for any x of smaller absolute value; that is, if a circle of radius X be drawn around the origin in the complex plane for x or about the point α in the complex plane for z, the series (9) and (9′) respectively will converge absolutely for all complex numbers which lie within these circles.



Three cases should be distinguished. First the series may converge for any value x no matter how great its absolute value. The circle may then have an indefinitely large radius; the series converge for all values of x or z and the function defined by them is finite (whether real or complex) for all values of the argument. Such a function is called an *integral function* of the complex variable z or x. Secondly, the series may converge for no other value than $x = 0$ or $z = \alpha$ and therefore cannot define any function. Thirdly, there may be a definite largest value for the radius, say R, such that for any point within the respective circles of radius R the series converge and define a function, whereas for any point outside the circles the series diverge. The circle of radius R is called the *circle of convergence* of the series.

As the matter of the radius and circle of convergence is important, it will be well to go over the whole matter in detail. Consider the suite of numbers

$$|a_1|, \quad \sqrt[2]{|a_2|}, \quad \sqrt[3]{|a_3|}, \quad \cdots, \quad \sqrt[n]{|a_n|}, \quad \cdots.$$

Let them be imagined to be located as points with coördinates between 0 and $+\infty$ on a line. Three possibilities as to the distribution of the points arise. First they

may be unlimited above, that is, it may be possible to pick out from the suite a set of numbers which increase without limit. Secondly, the numbers may converge to the limit 0. Thirdly, neither of these suppositions is true and the numbers from 0 to $+\infty$ may be divided into two classes such that every number in the first class is less than an infinity of numbers of the suite, whereas any number of the second class is surpassed by only a finite number of the numbers in the suite. The two classes will then have a frontier number which will be represented by $1/R$ (see §§ 19 ff.).

In the first case no matter what x may be it is possible to pick out members from the suite such that the set $\sqrt[i]{|a_i|}, \sqrt[j]{|a_j|}, \sqrt[k]{|a_k|}, \cdots$, with $i < j < k \cdots$, increases without limit. Hence the set $\sqrt[i]{|a_i|}|x|, \sqrt[j]{|a_j|}|x|, \cdots$ will increase without limit; the terms $a_ix^i\ a_jx^j, \cdots$ of the series (9′) do not approach 0 as their limit, and the series diverges for all values of x other than 0. In the second case the series converges for any value of x. For let ϵ be any number less than $1/|x|$. It is possible to go so far in the suite that all subsequent numbers of it shall be less than this assigned ϵ. Then

$$|a_{n+p}x^{n+p}| < \epsilon^{n+p}|x|^{n+p} \quad \text{and} \quad \epsilon^n|x|^n + \epsilon^{n+1}|x|^{n+1} + \cdots, \qquad \epsilon|x| < 1,$$

serves as a comparison series to insure the absolute convergence of (9′). In the third case the series converges for any x such that $|x| < R$ but diverges for any x such that $|x| > R$. For if $|x| < R$, take $\epsilon < R - |x|$ so that $|x| < R - \epsilon$. Now proceed in the suite so far that all the subsequent numbers shall be less than $1/(R-\epsilon)$, which is greater than $1/R$. Then

$$|a_{n+p}x^{n+p}| < \frac{|x|^{n+p}}{(R-\epsilon)^{n+p}} < 1, \quad \text{and} \quad \sum_0^\infty \frac{|x|^{n+p}}{(R-\epsilon)^{n+p}}$$

will do as a comparison series. If $|x| > R$, it is easy to show the terms of (9′) do not approach the limit 0.

Let a circle of radius r less than R be drawn concentric with the circle of convergence. Then *within the circle of radius $r < R$ the power series* (9′) *converges uniformly and defines a continuous function: the integral of the function may be had by integrating the series term by term,*

$$\Phi(x) = \int_0^x \phi(x)\,dx = a_0x + \frac{1}{2}a_1x^2 + \frac{1}{3}a_2x^3 + \cdots + \frac{1}{n}a_{n-1}x^n + \cdots;$$

and the series of derivatives converges uniformly and represents the derivative of the function,

$$\phi'(x) = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1} + \cdots.$$

To prove these theorems it is merely necessary to set up an M-series for the series itself and for the series of derivatives. Let X be any number between r and R. Then

$$|a_0| + |a_1|X + |a_2|X^2 + \cdots + |a_n|X^n + \cdots \tag{10}$$

converges because $X < R$; and furthermore $|a_n x^n| < |a_n| X^n$ holds for any x such that $|x| < X$, that is, for all points within and on the circle of radius r. Moreover as $|x| < X$,

$$|na_n x^{n-1}| = |a_n| \frac{n}{X}\left(\frac{|x|}{X}\right)^{n-1} X^n < |a_n| X^n$$

holds for sufficiently large values of n and for any x such that $|x| \leqq r$. Hence (10) serves as an M-series for the given series and the series of derivatives; and the theorems are proved. It should be noticed that it is incorrect to say that the convergence is uniform over the circle of radius R, although the statement is true of any circle within that circle no matter how small $R - r$. For an apparently slight but none the less important extension to include, in some cases, some points upon the circle of convergence see Ex. 5.

An immediate corollary of the above theorems is that *any power series* (9) *in the complex variable which converges for other values than* $z = a$, *and hence has a finite circle of convergence or converges all over the complex plane, defines an analytic function* $f(z)$ *of* z *in the sense of* §§ 73, 126; for the series is differentiable within any circle within the circle of convergence and thus the function has a definite finite and continuous derivative.

167. It is now possible to extend Taylor's and Maclaurin's Formulas, which developed a function of a real variable x into a polynomial plus a remainder, to *infinite series* known as Taylor's and Maclaurin's Series, which express the function as a power series, provided the remainder after n terms converges uniformly toward 0 as n becomes infinite. It will be sufficient to treat one case. Let

$$f(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \cdots + \frac{1}{(n-1)!}f^{(n-1)}(0)x^{n-1} + R_n,$$

$$R_n = \frac{x^n}{n!}f^{(n)}(\theta x) = \frac{x^n(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(\theta x) = \frac{1}{(n-1)!}\int_0^x t^{n-1}f^{(n)}(x-t)\,dt,$$

$$\lim_{n=\infty} R_n(x) = 0 \text{ uniformly in some interval } -h \leqq x \leqq h,$$

where the first line is Maclaurin's Formula, the second gives differnet forms of the remainder, and the third expresses the condition that the remainder converges to 0. Then the series

$$f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2$$
$$+ \cdots + \frac{1}{(n-1)!}f^{(n-1)}(0)x^{n-1} + \frac{1}{n!}f^{(n)}(0)x^n + \cdots \qquad (11)$$

converges to the value $f(x)$ for any x in the interval. The proof consists merely in noting that $f(x) - R_n(x) = S_n(x)$ is the sum of the first n terms of the series and that $|R_n(x)| < \epsilon$.

In the case of the exponential function e^x the nth derivative is e^x, and the remainder, taken in the first form, becomes

$$R_n(x) = \frac{1}{n!} e^{\theta x} x^n, \qquad |R_n(x)| < \frac{1}{n!} e^h h^n, \qquad |x| \leqq h.$$

As n becomes infinite, R_n clearly approaches zero no matter what the value of h; and

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots$$

is the infinite series for the exponential function. The series converges for all values of x real or complex and may be taken as the definition of e^x for complex values. This definition may be shown to coincide with that obtained otherwise (§ 74).

For the expansion of $(1 + x)^m$ the remainder may be taken in the second form.

$$R_n(x) = \frac{m(m-1)\cdots(m-n+1)}{1 \cdot 2 \cdots (n-1)} x^n \left(\frac{1-\theta}{1+\theta x}\right)^{n-1} (1 + \theta x)^{m-1},$$

$$|R_n(x)| < \left|\frac{m(m-1)\cdots(m-n+1)}{1 \cdot 2 \cdots (n-1)}\right| h^n (1+h)^{m-1}, \qquad h < 1.$$

Hence when $h < 1$ the limit of $R_n(x)$ is zero and the infinite expansion

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{2!} x^2 + \frac{m(m-1)(m-2)}{3!} x^3 + \cdots$$

is valid for $(1 + x)^m$ for all values of x numerically less than unity.

If in the binomial expansion x be replaced by $-x^2$ and m by $-\frac{1}{2}$,

$$\frac{1}{\sqrt{1-x^2}} = 1 + \frac{1}{2} x^2 + \frac{1 \cdot 3}{2 \cdot 4} x^4 + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} x^6 + \frac{1 \cdot 3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6 \cdot 8} x^8 + \cdots.$$

This series converges for all values of x numerically less than 1, and hence converges uniformly whenever $|x| \leqq h < 1$. It may therefore be integrated term by term.

$$\sin^{-1} x = x + \frac{1}{2}\frac{x^3}{3} + \frac{1 \cdot 3}{2 \cdot 4}\frac{x^5}{5} + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\frac{x^7}{7} + \frac{1 \cdot 3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6 \cdot 8}\frac{x^9}{9} + \cdots.$$

This series is valid for all values of x numerically less than unity. The series also converges for $x = \pm 1$, and hence by Ex. 5 is uniformly convergent when $-1 \leqq x \leqq 1$.

But Taylor's and Maclaurin's series may also be extended directly to functions $f(z)$ of a complex variable. If $f(z)$ is single valued and has a definite continuous derivative $f'(z)$ at every point of a region and on the boundary, the expansion

$$f(z) = f(a) + f'(a)(z-a) + \cdots + f^{(n-1)}(a)\frac{(z-a)^{n-1}}{(n-1)!} + R_n$$

has been established (§ 126) with the remainder in the form

$$|R_n(z)| = \left|\frac{(z-a)^n}{2\pi} \int_C \frac{f(t)\,dt}{(t-a)^n (t-z)}\right| \leqq \frac{1}{2\pi}\frac{r^n}{\rho^n}\frac{ML}{\rho - r}$$

for all points z within the circle of radius r (Ex. 7, p. 306). As n becomes infinite, R_n approaches zero uniformly, and hence the infinite series

$$f(z) = f(a) + f'(a)(z-a) + \cdots + f^{(n)}(a)\frac{(z-a)^n}{n!} + \cdots \tag{12}$$

is valid at all points within the circle of radius r and upon its circumference. The expansion is therefore convergent and valid for any z actually within the circle of radius ρ.

Even for real expansions (11) the significance of this result is great because, except in the simplest cases, it is impossible to compute $f^{(n)}(x)$ and establish the convergence of Taylor's series for real variables. The result just found shows that if the values of the function be considered for complex values z in addition to real values x, the circle of convergence will extend out to the nearest point where the conditions imposed on $f(z)$ break down, that is, to the nearest point at which $f(z)$ becomes infinite or otherwise ceases to have a definite continuous derivative $f'(z)$. For example, there is nothing in the behavior of the function

$$(1+x^2)^{-1} = 1 - x^2 + x^4 - x^6 + x^8 - \cdots,$$

as far as real values are concerned, which should indicate why the expansion holds only when $|x| < 1$; but in the complex domain the function $(1+z^2)^{-1}$ becomes infinite at $z = \pm i$, and hence the greatest circle about $z = 0$ in which the series could be expected to converge has a unit radius. Hence by considering $(1+z^2)^{-1}$ for complex values, it can be predicted without the examination of the nth derivative that the Maclaurin development of $(1+x^2)^{-1}$ will converge when and only when x is a proper fraction.

EXERCISES

1. (α) Does $x + x(1-x) + x(1-x)^2 + \cdots$ converge uniformly when $0 \leqq x \leqq 1$?

(β) Does the series $(1+k)^{\frac{1}{k}} = 1 + 1 + \frac{1-k}{2!} + \frac{(1-k)(1-2k)}{3!} + \cdots$ converge uniformly for small values of k? Can the derivation of the limit e of § 4 thus be made rigorous and the value be found by setting $k = 0$ in the series?

2. Test these series for uniform convergence; also the series of derivatives:

(α) $1 + x\sin\theta + x^2\sin 2\theta + x^3 \sin 3\theta + \cdots,\qquad |x| \leqq X < 1.$

(β) $1 + \frac{\sin x}{1^{2\frac{1}{2}}} + \frac{\sin^2 x}{2^{2\frac{1}{2}}} + \frac{\sin^3 x}{3^{2\frac{1}{2}}} + \frac{\sin^4 x}{4^{2\frac{1}{2}}} + \cdots,\qquad |x| \leqq X < \infty,$

(γ) $\frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \frac{1}{3}\left(\frac{x-1}{x}\right)^3 + \cdots,\qquad \frac{1}{2} < \gamma \leqq x \leqq X < \infty,$

(δ) $\frac{x-1}{x+1} + \frac{1}{3}\left(\frac{x-1}{x+1}\right)^3 + \frac{1}{5}\left(\frac{x-1}{x+1}\right)^5 + \cdots,\qquad 0 < \gamma \leqq x \leqq X < \infty.$

(ϵ) Consider complex as well as real values of the variable.

3. Determine the radius of convergence and draw the circle. Note that in practice the test ratio is more convenient than the theoretical method of the text:

(α) $x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots$. $\quad(\beta)$ $x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots$,

(γ) $\frac{1}{a}\left[1 + \frac{bx}{a} + \frac{b^2x^2}{a^2} + \frac{b^3x^3}{a^3} + \cdots\right]$. $\quad(\delta)$ $1 - x^2 + \frac{x^4}{2\,!} - \frac{x^6}{3\,!} + \frac{x^8}{4\,!} - \cdots$,

(ϵ) $\frac{1}{1}x - (\frac{1}{1} + \frac{1}{2})x^2 + (\frac{1}{1} + \frac{1}{2} + \frac{1}{3})x^3 - (\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4})x^4 + \cdots$,

(ζ) $1 - \frac{3^2+3}{4\cdot 2\,!}x^2 + \frac{3^4+3}{4\cdot 4\,!}x^4 - \frac{3^6+3}{4\cdot 6\,!}x^6 + \cdots$,

(η) $1 - x + x^4 - x^5 + x^8 - x^9 + x^{12} - x^{13} + \cdots$,

(θ) $(x-1)^1 - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \cdots$,

(ι) $x - \frac{(m-1)(m+2)}{3\,!}x^3 + \frac{(m-1)(m-3)(m+2)(m+4)}{5\,!}x^5 - \cdots$.

(κ) $1 - \frac{x^2}{2^2(m+1)} + \frac{x^4}{2^4\cdot 2\,!\,(m+1)(m+2)} - \frac{x^6}{2^6\cdot 3\,!\,(m+1)(m+2)(m+3)} + \cdots$.

(λ) $\frac{x^2}{2^2} - \frac{x^4}{2^4(2\,!)^2}\left(\frac{1}{1} + \frac{1}{2}\right) + \frac{x^6}{2^6(3\,!)^2}\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3}\right) - \frac{x^8}{2^8(4\,!)^2}\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) + \cdots$.

(μ) $1 + \frac{\alpha\beta}{1\cdot\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot 2\cdot\gamma(\gamma+1)}x^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot 2\cdot 3\cdot\gamma(\gamma+1)(\gamma+2)}x^3 + \cdots$.

4. Establish the Maclaurin expansions for the elementary functions:

(α) $\log(1-x)$, $\quad(\beta)$ $\sin x$, $\quad(\gamma)$ $\cos x$, $\quad(\delta)$ $\cosh x$,

(ϵ) a^x, $\quad(\zeta)$ $\tan^{-1}x$. $\quad(\eta)$ $\sinh^{-1}x$, $\quad(\theta)$ $\tanh^{-1}x$.

5. *Abel's Theorem.* If the infinite series $a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$ converges for the value X, it converges uniformly in the interval $0 \leqq x \leqq X$. Prove this by showing that (see Exs. 17–19, p. 428)

$$|R_n(x)| = |a_nx^n + a_{n+1}x^{n+1} + \cdots| < \left(\frac{x}{X}\right)^n |a_nX^n + \cdots + a_{n+p}X^{n+p}|,$$

when p is rightly chosen. Apply this to extending the interval over which the series is uniformly convergent to extreme values of the interval of convergence wherever possible in Exs. 4 (α), (ζ), (θ).

6. Examine sundry of the series of Ex. 3 in regard to their convergence at extreme points of the interval of convergence or at various other points of the circumference of their circle of convergence. Note the significance in view of Ex. 5.

7. Show that $f(x) = e^{-\frac{1}{x^2}}$, $f(0) = 0$, cannot be expanded into an infinite Maclaurin series by showing that $R_n = e^{-\frac{1}{x^2}}$, and hence that R_n does not converge uniformly toward 0 (see Ex. 9, p. 66). Show this also from the consideration of complex values of x.

8. From the consideration of complex values determine the interval of convergence of the Maclaurin series for

(α) $\tan x = \frac{\sin x}{\cos x}$, $\quad(\beta)$ $\frac{x}{e^x - 1}$, $\quad(\gamma)$ $\tanh x$. $\quad(\delta)$ $\log(1 + e^x)$.

9. Show that if two similar infinite power series represent the same function in any interval the coefficients in the series must be equal (cf. § 32).

10. From $1 + 2r\cos x + r^2 = (1 + re^{ix})(1 + re^{-ix}) = r^2\left(1 + \frac{e^{ix}}{r}\right)\left(1 + \frac{e^{-ix}}{r}\right)$

prove $\quad \log(1 + 2r\cos x + r^2) = 2\left(r\cos x - \frac{r^2}{2}\cos 2x + \frac{r^3}{3}\cos 3x - \cdots\right)$,

$$\int_0^x \log(1 + 2r\cos x + r^2)\,dx = 2\left(r\sin x - \frac{r^2}{2^2}\sin 2x + \frac{r^3}{3^2}\sin 3x - \cdots\right); \qquad r < 1$$

and $\quad \log(1 + 2r\cos x + r^2) = 2\log r + 2\left(\frac{\cos x}{r} - \frac{\cos 2x}{2r^2} + \frac{\cos 3x}{3r^3} - \cdots\right)$,

$$\int_0^x \log(1 + 2r\cos x + r^2)\,dx = 2x\log r + 2\left(\frac{\sin x}{r} - \frac{\sin 2x}{2^2r^2} + \frac{\sin 3x}{3^2r^3} - \cdots\right); \qquad r > 1$$

$$\int_0^x \log(1 + \sin\alpha\cos x)\,dx = 2x\log\cos\frac{\alpha}{2} + 2\left(\tan\frac{\alpha}{2}\sin x - \tan^2\frac{\alpha}{2}\frac{\sin 2x}{2^2} + \cdots\right).$$

11. Prove $\displaystyle\int_0^1 \frac{dx}{\sqrt{1+x^4}} = 1 - \frac{1}{2\cdot 5} + \frac{1\cdot 3}{2\cdot 4\cdot 9} - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 13} + \cdots = \int_1^\infty \frac{dx}{\sqrt{1+x^4}}$.

12. Evaluate these integrals by expansion into series (see Ex. 23, p. 452)

$(\alpha)\quad \displaystyle\int_0^\infty \frac{e^{-qx}\sin rx}{x}\,dx = \frac{r}{q} - \frac{1}{3}\left(\frac{r}{q}\right)^3 + \frac{1}{5}\left(\frac{r}{q}\right)^5 - \cdots = \tan^{-1}\frac{r}{q}$,

$(\beta)\quad \displaystyle\int_0^\pi \frac{\log(1 + k\cos x)}{\cos x}\,dx = \pi\sin^{-1}k, \qquad (\gamma)\quad \int_0^\pi \frac{x\sin x}{1+\cos^2 x}\,dx = \frac{\pi^2}{4}$,

$(\delta)\quad \displaystyle\int_0^\infty e^{-a^2x^2}\cos 2\beta x\,dx = \frac{\sqrt{\pi}}{2a}e^{-\left(\frac{\beta}{a}\right)^2}, \qquad (\epsilon)\quad \int_0^\pi \log(1 + 2r\cos x + r^2)\,dx.$

13. By formal multiplication (§ 168) show that

$$\frac{1-\alpha^2}{1 - 2\alpha\cos x + \alpha^2} = 1 + 2\alpha\cos x + 2\alpha^2\cos 2x + \cdots,$$

$$\frac{\alpha\sin x}{1 - 2\alpha\cos x + \alpha^2} = \alpha\sin x + \alpha^2\sin 2x + \cdots.$$

14. Evaluate, by use of Ex. 13, these definite integrals, m an integer:

$(\alpha)\quad \displaystyle\int_0^\pi \frac{\cos mx\,dx}{1 - 2\alpha\cos x + \alpha^2} = \frac{\pi\alpha^m}{1-\alpha^2}, \qquad (\beta)\quad \int_0^\pi \frac{x\sin x\,dx}{1 - 2\alpha\cos x + \alpha^2} = \frac{\pi}{\alpha}\log(1+\alpha).$

$$(\gamma)\quad \int_0^\pi \frac{\sin x\sin mx\,dx}{1 - 2\alpha\cos x + \alpha^2} = \frac{\pi}{2}\alpha^{m-1},$$

$$(\delta)\quad \int_0^\pi \frac{\sin^2 x\,dx}{(1 - 2\alpha\cos x + \alpha^2)(1 - 2\beta\cos x + \beta^2)}.$$

15. In Ex. 14 (γ) let $\alpha = 1 - h/m$ and $x = z/m$. Obtain by a limiting process, and by a similar method exercised upon Ex. 14 (α):

$$\int_0^\infty \frac{z\sin z\,dz}{h^2 + z^2} = \frac{\pi}{2}e^{-h}, \qquad \int_0^\infty \frac{\cos z\,dz}{h^2+z^2} = \frac{\pi}{2}e^{-h}.$$

Can the use of these limiting processes be readily justified?

16. Let h and x be less than 1. Assume the expansion

$$f(x, h) = \frac{1}{\sqrt{1 - 2xh + h^2}} = 1 + hP_1(x) + h^2P_2(x) + \cdots + h^nP_n(x) + \cdots.$$

Obtain therefrom the following expansions by differentiation:

$$\frac{1}{h}f'_x = \frac{1}{(1 - 2xh + h^2)^{\frac{3}{2}}} = P'_1 + hP'_2 + h^2P'_3 + \cdots + h^{n-1}P'_n + \cdots,$$

$$f'_h = \frac{x - h}{(1 - 2xh + h^2)^{\frac{3}{2}}} = P_1 + 2hP_2 + 3h^2P_3 + \cdots + nh^{n-1}P_n + \cdots.$$

Hence establish the given identities and consequent relations:

$$\begin{aligned}
\frac{x-h}{h}f'_x &= xP'_1 + h(xP'_2 - P'_1) + \cdots + h^{n-1}(xP'_n - P'_{n-1}) + \cdots = \\
f'_h &= P_1 + h(2P_2) + \cdots + h^{n-1}(nP_n) + \cdots, \\
\frac{(1+h^2)}{h}f'_x - f &= -1 + P'_1 + h(P'_2 - P_1) + \cdots + h^n(P'_{n+1} + P'_{n-1} - P_n) + \cdots = \\
2xhf &= h(2x) + \cdots + h^n(2xP_{n-1}).
\end{aligned}$$

Or $\qquad nP_n = xP'_n - P'_{n-1} \qquad$ and $\quad P'_{n+1} + P'_{n-1} - P_n = 2xP'_n.$

Hence $\quad xP'_n = P'_{n+1} - (n+1)P_n \quad$ and $\quad (x^2 - 1)P'_n = n(xP_n - P_{n-1}).$

Compare the results with Exs. 13 and 17, p. 252, to identify the functions with the Legendre polynomials. Write

$$\frac{1}{(1 - 2xh + h^2)^{\frac{1}{2}}} = \frac{1}{(1 - 2h\cos\theta + h^2)^{\frac{1}{2}}} = \frac{1}{(1 - he^{i\theta})^{\frac{1}{2}}(1 - he^{-i\theta})^{\frac{1}{2}}}$$
$$= \left(1 + \frac{1}{2}he^{i\theta} + \frac{1\cdot 3}{2\cdot 4}h^2e^{2i\theta} + \cdots\right)\left(1 + \frac{1}{2}he^{-i\theta} + \frac{1\cdot 3}{2\cdot 4}h^2e^{-2i\theta} + \cdots\right),$$

and show $P_n(\cos\theta) = 2\,\dfrac{1\cdot 3\cdots(2n-1)}{2\cdot 4\cdots 2n}\left\{\cos n\theta + \dfrac{1\cdot n}{1\cdot(2n-1)}\cos(n-2)\theta + \cdots\right\}.$

168. Manipulation of series. *If an infinite series*

$$S = u_0 + u_1 + u_2 + \cdots + u_{n-1} + u_n + \cdots \tag{13}$$

converges, the series obtained by grouping the terms in parentheses without altering their order will also converge. Let

$$S' = U_0 + U_1 + \cdots + U_{n'-1} + U_{n'} + \cdots \tag{13'}$$

and $$S'_1,\ S'_2,\ \cdots,\ S'_{n'},\ \cdots$$

be the new series and the sums of its first n' terms. These sums are merely particular ones of the set $S_1, S_2, \cdots, S_n, \cdots$, and as $n' < n$ it follows that n becomes infinite when n' does if n be so chosen that $S_n = S'_{n'}$. As S_n approaches a limit, $S'_{n'}$ must approach the same limit. As a corollary it appears that if the series obtained by removing parentheses in a given series converges, the value of the series is not affected by removing the parentheses.

If two convergent infinite series be given as

$$S = u_0 + u_1 + \cdots, \quad \textit{and} \quad T = v_0 + v_1 + \cdots,$$

then $$(\lambda u_0 + \mu v_0) + (\lambda u_1 + \mu v_1) + \cdots$$

will converge to the limit $\lambda S + \mu T$, and will converge absolutely provided both the given series converge absolutely. The proof is left to the reader.

If a given series converges absolutely, the series formed by rearranging the terms in any order without omitting any terms will converge to the same value. Let the two arrangements be

$$S = u_0 + u_1 + u_2 + \cdots + u_{n-1} + u_n + \cdots$$

and $$S' = u_{0'} + u_{1'} + u_{2'} + \cdots + u_{n'-1} + u_{n'} + \cdots.$$

As S converges absolutely, n may be taken so large that

$$|u_n| + |u_{n+1}| + \cdots < \epsilon;$$

and as the terms in S' are identical with those in S except for their order, n' may be taken so large that $S'_{n'}$ shall contain all the terms in S_n. The other terms in $S'_{n'}$ will be found among the terms $u_n, u_{n+1}, \cdots$. Hence $$|S'_{n'} - S_n| < |u_n| + |u_{n+1}| + \cdots < \epsilon.$$

As $|S - S_n| < \epsilon$, it follows that $|S - S'_{n'}| < 2\epsilon$. Hence $S'_{n'}$ approaches S as a limit when n' becomes infinite. It may easily be shown that S' also converges absolutely.

The theorem is still true if the rearrangement of S is into a series some of whose terms are themselves infinite series of terms selected from S. Thus let $$S' = U_0 + U_1 + U_2 + \cdots + U_{n'-1} + U_{n'} + \cdots,$$

where U_i may be any aggregate of terms selected from S. If U_i be an infinite series of terms selected from S, as

$$U_i = u_{i0} + u_{i1} + u_{i2} + \cdots + u_{in} + \cdots,$$

the absolute convergence of U_i follows from that of S (cf. Ex. 22 below). It is possible to take n' so large that every term in S_n shall occur in one of the terms $U_0, U_1, \cdots, U_{n'-1}$. Then if from

$$S - U_0 - U_1 - \cdots - U_{n'-1} \tag{14}$$

there be canceled all the terms of S_n, the terms which remain will be found among $u_n, u_{n+1}, \cdots$, and (14) will be less than ϵ. Hence as n' becomes infinite, the difference (14) approaches zero as a limit and the theorem is proved that

$$S = U_0 + U_1 + \cdots + U_{n'-1} + U_{n'} + \cdots = S'.$$

If a series of real terms is convergent, but not absolutely, the number of positive and the number of negative terms is infinite, the series of positive terms and the series of negative terms diverge, and the given series may be so rearranged as to comport itself in any desired manner. That the number of terms of each sign cannot be finite follows from the fact that if it were, it would be possible to go so far in the series that all subsequent terms would have the same sign and the series would therefore converge absolutely if at all. Consider next the sum $S_n = P_l - N_m$, $l + m = n$, of n terms of the series, where P_l is the sum of the positive terms and N_m that of the negative terms. If both P_l and N_m converged, then $P_l + N_m$ would also converge and the series would converge absolutely; if only one of the sums P_l or N_m diverged, then S would diverge. Hence both sums must diverge. The series may now be rearranged to approach any desired limit, to become positively or negatively infinite, or to oscillate as desired. For suppose an arrangement to approach L as a limit were desired. First take enough positive terms to make the sum exceed L, then enough negative terms to make it less than L, then enough positive terms to bring it again in excess of L, and so on. But as the given series converges, its terms approach 0 as a limit; and as the new arrangement gives a sum which never differs from L by more than the last term in it, the difference between the sum and L is approaching 0 and L is the limit of the sum. In a similar way it could be shown that an arrangement which would comport itself in any of the other ways mentioned would be possible.

If two absolutely convergent series be multiplied, as

$$S = u_0 + u_1 + u_2 + \cdots + u_n + \cdots,$$

$$T = v_0 + v_1 + v_2 + \cdots + v_n + \cdots,$$

and

$$\begin{aligned} W = {} & u_0v_0 + u_1v_0 + u_2v_0 + \cdots + u_nv_0 + \cdots \\ & + u_0v_1 + u_1v_1 + u_2v_1 + \cdots + u_nv_1 + \cdots \\ & + \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\ & + u_0v_n + u_1v_n + u_2v_n + \cdots + u_nv_n + \cdots \\ & + \quad . \quad . \quad . \quad . \quad . \quad . \quad . \end{aligned}$$

and if the terms in W be arranged in a simple series as

$$u_0v_0 + (u_1v_0 + u_1v_1 + u_0v_1) + (u_2v_0 + u_2v_1 + u_2v_2 + u_1v_2 + u_0v_2) + \cdots$$

or in any other manner whatsoever, the series is absolutely convergent and converges to the value of the product ST.

In the particular arrangement above, S_1T_1, S_2T_2, S_nT_n is the sum of the first, the first two, the first n terms of the series of parentheses. As $\lim S_nT_n = ST$, the series of parentheses converges to ST. As S and T are absolutely convergent the same reasoning could be applied to the series of absolute values and

$$|u_0||v_0| + |u_1||v_0| + |u_1||v_1| + |u_0||v_1| + |u_2||v_0| + \cdots$$

would be seen to converge. Hence the convergence of the series

$$u_0v_0 + u_1v_0 + u_1v_1 + u_0v_1 + u_2v_0 + u_2v_1 + u_2v_2 + u_1v_2 + u_0v_2 + \cdots$$

is absolute and to the value ST when the parentheses are omitted. Moreover, any other arrangement, such in particular as

$$u_0v_0 + (u_1v_0 + u_0v_1) + (u_2v_0 + u_1v_1 + u_0v_2) + \cdots,$$

would give a series converging absolutely to ST.

The equivalence of a function and its Taylor or Maclaurin infinite series (wherever the series converges) lends importance to the operations of multiplication, division, and so on, which may be performed on the series. Thus if

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots, \qquad |x| < R_1,$$
$$g(x) = b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots, \qquad |x| < R_2,$$

the multiplication may be performed and the series arranged as

$$f(x)g(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \cdots$$

according to ascending powers of x whenever x is numerically less than the smaller of the two radii of convergence R_1, R_2, because both series will then converge absolutely. Moreover, Ex. 5 above shows that this form of the product may still be applied at the extremities of its interval of convergence for real values of x provided the series converges for those values.

As an example in the multiplication of series let the product $\sin x \cos x$ be found.

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots, \qquad \cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \cdots.$$

The product will contain only odd powers of x. The first few terms are

$$1x - \left(\frac{1}{3!} + \frac{1}{2!}\right)x^3 + \left(\frac{1}{5!} + \frac{1}{3!\,2!} + \frac{1}{4!}\right)x^5 - \left(\frac{1}{7!} + \frac{1}{5!\,2!} + \frac{1}{3!\,4!} + \frac{1}{6!}\right)x^7 + \cdots.$$

The law of formation of the coefficients gives as the coefficient of x^{2k+1}

$$(-1)^k\left[\frac{1}{(2k+1)!} + \frac{1}{(2k-1)!\,2!} + \frac{1}{(2k-3)!\,4!} + \cdots + \frac{1}{3!\,(2k-2)!} + \frac{1}{(2k)!}\right] =$$
$$\frac{(-1)^k}{(2k+1)!}\left[1 + \frac{(2k+1)2k}{2!} + \frac{(2k+1)(2k)(2k-1)(2k-2)}{4!} + \cdots + \frac{(2k+1)}{1!}\right].$$

But $\quad 2^{2k+1} = (1+1)^{2k+1} = 1 + (2k+1) + \dfrac{(2k+1)2k}{2!} + \cdots + (2k+1) + 1.$

Hence it is seen that the coefficient of x^{2k+1} takes every other term in this symmetrical sum of an even number of terms and must therefore be equal to half the sum. The product may then be written as the series

$$\sin x \cos x = \frac{1}{2}\left[2x - \frac{(2x)^3}{3!} + \frac{(2x)^5}{5!} - \cdots\right] = \frac{1}{2}\sin 2x.$$

169. If a function $f(x)$ be expanded into a power series

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots, \qquad |x| < R, \tag{15}$$

and if $x = \alpha$ is any point within the circle of convergence, *it may be desired to transform the series into one which proceeds according to powers of $(x - \alpha)$ and converges in a circle about the point $x = \alpha$.* Let $t = x - \alpha$. Then $x = \alpha + t$ and hence

$$x^2 = \alpha^2 + 2\,\alpha t + t^2, \qquad x^3 = \alpha^3 + 3\,\alpha^2 t + 3\,\alpha t^2 + t^3, \qquad \cdots,$$

$$f(x) = a_0 + a_1(\alpha + t) + a^2(\alpha^2 + 2\,\alpha t + t^2) + \cdots. \tag{15'}$$

Since $|\alpha| < R$, the relation $|\alpha| + |t| < R$ will hold for small values of t, and the series (15′) will converge for $x = |\alpha| + |t|$. Since

$$a_0 + a_1(|\alpha| + |t|) + a_2(|\alpha|^2 + 2\,|\alpha||t| + |t|^2) + \cdots$$

is absolutely convergent for small values of t, the parentheses in (15′) may be removed and the terms collected as

$$f(x) = \phi(t) = (a_0 + a_1\alpha + a_2\alpha^2 + a_3\alpha^3 + \cdots) + (a_1 + 2\,a_2\alpha + 3\,a_3\alpha^2 + \cdots)t$$
$$+ (a_2 + 3\,a_3\alpha + \cdots)t^2 + (a_3 + \cdots)t^3 + \cdots,$$

or
$$f(x) = \phi(x - \alpha) = A_0 + A_1(x - \alpha) + A_2(x - \alpha)^2 + A_3(x - \alpha)^3 + \cdots, \tag{16}$$

where $A_0, A_1, A_2, \cdots$ are infinite series; in fact

$$A_0 = f(\alpha), \qquad A_1 = f'(\alpha), \qquad A_2 = \frac{1}{2\,!}f''(\alpha), \qquad A_3 = \frac{1}{3\,!}f'''(\alpha), \cdots.$$

The series (16) in $x - \alpha$ will surely converge within a circle of radius $R - |\alpha|$ about $x = \alpha$; but it may converge in a larger circle. As a matter of fact it will converge within the largest circle whose center is at α and within which the function has a definite continuous derivative. Thus Maclaurin's expansion for $(1 + x^2)^{-1}$ has a unit radius of convergence; but the expansion about $x = \frac{1}{2}$ into powers of $x - \frac{1}{2}$ will have a radius of convergence equal to $\frac{1}{2}\sqrt{5}$, which is the distance from $x = \frac{1}{2}$ to either of the points $x = \pm i$. If the function had originally been defined by its development about $x = 0$, the definition would have been valid only over the unit circle. The new development about $x = \frac{1}{2}$ will therefore extend the definition to a considerable region outside the original domain, and by repeating the process the region of definition may be extended further. As the function is at each step defined by a power series, it remains analytic. This process of extending the definition of a function is called *analytic continuation.*

Consider *the expansion of a function of a function.* Let

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \qquad |x| < R_1,$$
$$x = \phi(y) = b_0 + b_1 y + b_2 y^2 + b_3 y^3 + \cdots, \qquad |y| < R_2,$$

and let $|b_0| < R_1$ so that, for sufficiently small values of y, the point x will still lie within the circle R_1. By the theorem on multiplication, the series for x may be squared, cubed, $\cdots$, and the series for x^2, x^3, $\cdots$ may be arranged according to powers of y. These results may then be substituted in the series for $f(x)$ and the result may be ordered according to powers of y. Hence the expansion for $f[\phi(y)]$ is obtained. That the expansion is valid at least for small values of y may be seen by considering

$$|a_0| + |a_1|\xi + |a_2|\xi^2 + |a_3|\xi^3 + \cdots, \qquad \xi < R_1,$$
$$\xi = |b_0| + |b_1||y| + |b_2||y|^2 + \cdots, \qquad |y| \text{ small},$$

which are series of positive terms. The radius of convergence of the series for $f[\phi(y)]$ may be found by discussing that function.

For example consider the problem of expanding $e^{\cos x}$ to five terms.

$$e^y = 1 + y + \tfrac{1}{2} y^2 + \tfrac{1}{6} y^3 + \tfrac{1}{24} y^4 + \cdots, \qquad y = \cos x = 1 - \tfrac{1}{2} x^2 + \tfrac{1}{24} x^4 + \cdots,$$

$$y^2 = 1 - x^2 + \tfrac{1}{3} x^4 - \cdots, \quad y^3 = 1 - \tfrac{3}{2} x^2 + \tfrac{7}{8} x^4 - \cdots, \quad y^4 = 1 - 2x^2 + 1\tfrac{2}{3} x^4 - \cdots,$$

$$\begin{aligned} e^y = 1 &+ (1 - \tfrac{1}{2} x^2 + \tfrac{1}{24} x^4 - \cdots) + \tfrac{1}{2}(1 - x^2 + \tfrac{1}{3} x^4 - \cdots) + \tfrac{1}{6}(1 - \tfrac{3}{2} x^2 + \tfrac{7}{8} x^4 - \cdots) \\ &+ \tfrac{1}{24}(1 - 2x^2 + 1\tfrac{2}{3} x^4 - \cdots) + \cdots \\ = (1 &+ 1 + \tfrac{1}{2} + \tfrac{1}{6} + \tfrac{1}{24} + \cdots) - (\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{12} + \cdots) x^2 \\ &+ (\tfrac{1}{2} + \tfrac{1}{6} + \tfrac{7}{48} + \tfrac{1}{9} + \cdots) x^4 + \cdots, \end{aligned}$$

$$e^y = e^{\cos x} = 2\tfrac{17}{24} - 1\tfrac{1}{3} x^2 + \tfrac{22}{27} x^4 - \cdots.$$

It should be noted that the coefficients in this series for $e^{\cos x}$ are really infinite series and the final values here given are only the approximate values found by taking the first few terms of each series. This will always be the case when $y = b_0 + b_1 x + \cdots$ begins with $b_0 \neq 0$; it is also true in the expansion about a new origin, as in a previous paragraph. In the latter case the difficulty cannot be avoided, but in the case of the expansion of a function of a function it is sometimes possible to make a preliminary change which materially simplifies the final result in that the coefficients become finite series. Thus here

$$e^{\cos x} = e^{1+z} = ee^z, \qquad z = \cos x - 1 = -\tfrac{1}{2} x^2 + \tfrac{1}{24} x^4 - \tfrac{1}{720} x^6 + \cdots,$$

$$z^2 = \tfrac{1}{4} x^4 - \tfrac{1}{24} x^6 + \cdots, \qquad z^3 = -\tfrac{1}{8} x^6 + \cdots, \qquad z^4, z^5, z^6 = 0 + \cdots,$$

$$e^z = 1 + (-\tfrac{1}{2} x^2 + \tfrac{1}{24} x^4 - \tfrac{1}{720} x^6 + \cdots) + \tfrac{1}{2}(\tfrac{1}{4} x^4 - \tfrac{1}{24} x^6 + \cdots) + \tfrac{1}{6}(-\tfrac{1}{8} x^6 + \cdots) + \cdots,$$

$$e^{\cos x} = ee^z = e(1 - \tfrac{1}{2} x^2 + \tfrac{1}{6} x^4 - \tfrac{31}{720} x^6 + \cdots).$$

The coefficients are now exact and the computation to x^6 turns out to be easier than to x^2 by the previous method; the advantage introduced by the change would be even greater if the expansion were to be carried several terms farther.

The quotient of two power series $f(x)$ by $g(x)$, if $g(0) \neq 0$, may be obtained by the ordinary algorism of division as

$$\frac{f(x)}{g(x)} = \frac{a_0 + a_1x + a_2x^2 + \cdots}{b_0 + b_1x + b_2x^2 + \cdots} = c_0 + c_1x + c_2x^2 + \cdots, \qquad b_0 \neq 0.$$

For in the first place as $g(0) \neq 0$, the quotient is analytic in the neighborhood of $x = 0$ and may be developed into a power series. It therefore merely remains to show that the coefficients $c_0, c_1, c_2, \cdots$ are those that would be obtained by division. Multiply

$$\begin{aligned}(a_0 + a_1x + a_2x^2 + \cdots) &= (c_0 + c_1x + c_2x^2 + \cdots)(b_0 + b_1x + b_2x^2 + \cdots)\\ &= b_0c_0 + (b_1c_0 + b_0c_1)x + (b_2c_0 + b_1c_1 + b_0c_2)x^2 + \cdots,\end{aligned}$$

and then equate coefficients of equal powers of x. Then

$$a_0 = b_0c_0, \qquad a_1 = b_1c_0 + b_0c_1, \qquad a_2 = b_2c_0 + b_1c_1 + b_0c_2, \cdots$$

is a set of equations to be solved for $c_0, c_1, c_2, \cdots$. The terms in $f(x)$ and $g(x)$ beyond x^n have no effect upon the values of $c_0, c_1, \cdots, c_n$, and hence these would be the same if $b_{n+1}, b_{n+2}, \cdots$ were replaced by $0, 0, \cdots$, and $a_{n+1}, a_{n+2}, \cdots, a_{2n}, a_{2n+1}, \cdots$ by such values $a'_{n+1}, a'_{n+2}, \cdots, a'_{2n}, 0, \cdots$ as would make the division come out even; the coefficients $c_0, c_1, \cdots, c_n$ are therefore precisely those obtained in dividing the series.

If y is developed into a power series in x as

$$y = f(x) = a_0 + a_1x + a_2x^2 + \cdots, \qquad a_1 \neq 0, \tag{17}$$

then x may be developed into a power series in $y - a_0$ as

$$x = f^{-1}(y - a_0) = b_1(y - a_0) + b_2(y - a_0)^2 + \cdots. \tag{18}$$

For since $a_1 \neq 0$, the function $f(x)$ has a nonvanishing derivative for $x = 0$ and hence the inverse function $f^{-1}(y - a_0)$ is analytic near $x = 0$ or $y = a_0$ and can be developed (p. 477). The method of undetermined coefficients may be used to find $b_1, b_2, \cdots$. This process of finding (18) from (17) is called the *reversion* of (17). For the actual work it is simpler to replace $(y - a_0)/a_1$ by t so that

$$t = x + a_2'x^2 + a_3'x^3 + a_4'x^4 + \cdots, \qquad a_i' = a_i/a_1,$$

and $$x = t + b_2't^2 + b_3't^3 + b_4't^4 + \cdots, \qquad b_i' = b_ia_1^i.$$

Let the assumed value of x be substituted in the series for t; rearrange the terms according to powers of t and equate the corresponding coefficients. Thus

$$\begin{aligned}t = t &+ (b_2' + a_2')t^2 + (b_3' + 2\,b_2'a_2' + a_3')t^3\\ &+ (b_4' + 2\,b_3'a_2' + b_2'^2a_2' + 3\,b_2'a_3' + a_4')t^4 + \cdots\end{aligned}$$

or $$b_2' = -a_2', \qquad b_3' = 2\,a_2'^2 - a_3', \qquad b_4' = -5\,a_2'^3 + 5\,a_2'a_3' - a_4', \cdots.$$

170. For some few purposes, which are tolerably important, *a formal operational method* of treating series is so useful as to be almost indispensable. If the series be taken in the form

$$1 + a_1 x + \frac{a_2}{2!}x^2 + \frac{a_3}{3!}x^3 + \cdots + \frac{a_n}{n!}x^n + \cdots,$$

with the factorials which occur in Maclaurin's development and with unity as the initial term, the series may be written as

$$e^{ax} = 1 + a^1 x + \frac{a^2}{2!}x^2 + \frac{a^3}{3!}x^3 + \cdots + \frac{a^n}{n!}x^n + \cdots,$$

provided that a^i be interpreted as the formal equivalent of a_i. The product of two series would then formally suggest

$$e^{ax}e^{bx} = e^{(a+b)x} = 1 + (a+b)^1 x + \frac{1}{2!}(a+b)^2 x^2 + \cdots, \tag{19}$$

and if the coefficients be transformed by setting $a^i b^j = a_i b_j$, then

$$\left(1 + a_1 x + \frac{a_2}{2!}x^2 + \cdots\right)\left(1 + b_1 x + \frac{b_2}{2!}x^2 + \cdots\right)$$
$$= 1 + (a_1 + b_1)x + \frac{a_2 + 2a_1 b_1 + b_2}{2!}x^2 + \cdots.$$

This as a matter of fact is the formula for the product of two series and hence justifies the suggestion contained in (19).

For example suppose that the development of

$$\frac{x}{e^x - 1} = 1 + B_1 x + \frac{B_2}{2!}x^2 + \frac{B_3}{3!}x^3 + \cdots$$

were desired. As the development begins with 1, the formal method may be applied and the result is found to be

$$\frac{x}{e^x - 1} = e^{Bx}, \qquad x = e^{(B+1)x} - e^{Bx}, \tag{20}$$

$$x = x + [(B+1)^2 - B^2]\frac{x^2}{2!} + [(B+1)^3 - B^3]\frac{x^3}{3!} + \cdots, \tag{21}$$

$$(B+1)^2 - B^2 = 0, \quad (B+1)^3 - B^3 = 0, \cdots, \quad (B+1)^k - B^k = 0, \cdots,$$

$$\text{or } 2B_1 + 1 = 0, \quad 3B_2 + 3B_1 + 1 = 0, \quad 4B_3 + 6B_2 + 4B_1 + 1 = 0, \cdots,$$

$$\text{or} \qquad B_1 = -\tfrac{1}{2}, \quad B_2 = \tfrac{1}{6}, \quad B_3 = 0, \quad B_4 = -\tfrac{1}{30}, \cdots.$$

The formal method leads to a set of equations from which the successive B's may quickly be determined. Note that

$$\frac{x}{e^x - 1} + \frac{x}{2} = \frac{x}{2}\frac{e^x + 1}{e^x - 1} = \frac{x}{2}\coth\frac{x}{2} = -\frac{x}{2}\coth\left(-\frac{x}{2}\right) \tag{22}$$

is an even function of x, and that consequently all the B's with odd indices except B_1 are zero. This will facilitate the calculation. The first eight even B's are respectively

$$\tfrac{1}{6}, \quad -\tfrac{1}{30}, \quad \tfrac{1}{42}, \quad -\tfrac{1}{30}, \quad \tfrac{5}{66}, \quad -\tfrac{691}{2730}, \quad \tfrac{7}{6}, \quad -\tfrac{3617}{510}. \tag{23}$$

The numbers B, or their absolute values, are called *the Bernoullian numbers*. An independent justification for the method of formal calculation may readily be given. For observe that $e^x e^{Bx} = e^{(B+1)x}$ of (20) is true when B is regarded as an independent variable. Hence if this identity be arranged according to powers of B, the coefficient of each power must vanish. It will therefore not disturb the identity if any numbers whatsoever are substituted for $B^1, B^2, B^3, \cdots$; the particular set $B_1, B_2, B_3, \cdots$ may therefore be substituted; the series may be rearranged according to powers of x, and the coefficients of like powers of x may be equated to 0, — as in (21) to get the desired equations.

If an infinite series be written without the factorials as

$$1 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n + \cdots,$$

a possible symbolic expression for the series is

$$\frac{1}{1 - ax} = 1 + a^1 x + a^2 x^2 + a^3 x^3 + \cdots, \qquad a^i = a_i.$$

If the substitution $y = x/(1 + x)$ or $x = y/(1 - y)$ be made,

$$\frac{1}{1 - ax} = \frac{1}{1 - a\dfrac{y}{1 - y}} = \frac{1 - y}{1 - (1 + a)y}. \tag{24}$$

Now if the left-hand and right-hand expressions be expanded and a be regarded as an independent variable restricted to values which make $|ax| < 1$, the series obtained will both converge absolutely and may be arranged according to powers of a. Corresponding coefficients will then be equal and the identity will therefore not be disturbed if a_i replaces a^i. Hence

$$1 + a_1 x + a_2 x^2 + \cdots = (1 - y)[1 + (1 + a)y + (1 + a)^2 y^2 + \cdots],$$

provided that both series converge absolutely for $a_i = a^i$. Then

$$\begin{aligned} 1 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots &= 1 + ay + a(1 + a)y^2 + a(1 + a)^2 y^3 + \cdots \\ &= 1 + a_1 y + (a_1 + a_2)y^2 + (a_1 + 2a_2 + a_3)y^3 + \cdots, \end{aligned}$$

or
$$\begin{aligned} a_1 x + a_2 x^2 + a_3 x^3 + \cdots &= a_1 y + (a_1 + a_2)y^2 \\ &\quad + (a_1 + 2a_2 + a_3)y^3 + \cdots. \end{aligned} \tag{25}$$

This transformation is known as *Euler's transformation.* Its great advantage for computation lies in the fact that sometimes the second series converges much more rapidly than the first. This is especially true when the coefficients of the first series are such as to make the coefficients in the new series small. Thus from (25)

$$\log(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \tfrac{1}{5}x^5 - \tfrac{1}{6}x^6 + \cdots$$
$$= y + \tfrac{1}{2}y^2 + \tfrac{1}{3}y^3 + \tfrac{1}{4}y^4 + \tfrac{1}{5}y^5 + \tfrac{1}{6}y^6 + \cdots.$$

To compute log 2 to three decimals from the first series would require several hundred terms; eight terms are enough with the second series. An additional advantage of the new series is that it may continue to converge after the original series has ceased to converge. In this case the two series can hardly be said to be equal; but the second series of course remains equal to the (continuation of the) function defined by the first. Thus log 3 may be computed to three decimals with about a dozen terms of the second series, but cannot be computed from the first.

EXERCISES

1. By the multiplication of series prove the following relations:

(α) $(1 + x + x^2 + x^3 + \cdots)^2 = (1 + 2x + 3x^2 + 4x^3 + \cdots) = (1-x)^{-2}$,

(β) $\cos^2 x + \sin^2 x = 1$, $\quad(\gamma)$ $e^x e^y = e^{x+y}$, $\quad(\delta)$ $2\sin^2 x = 1 - \cos 2x$.

2. Find the Maclaurin development to terms in x^6 for the functions:

(α) $e^x \cos x$, $\quad(\beta)$ $e^x \sin x$, $\quad(\gamma)$ $(1+x)\log(1+x)$, $\quad(\delta)$ $\cos x \sin^{-1} x$.

3. Group the terms of the expansion of $\cos x$ in two different ways to show that $\cos 1 > 0$ and $\cos 2 < 0$. Why does it then follow that $\cos\xi = 0$ where $1 < \xi < 2$?

4. Establish the developments (Peirce's Nos. 785–789) of the functions:

(α) $e^{\sin x}$, $\quad(\beta)$ $e^{\tan x}$, $\quad(\gamma)$ $e^{\sin^{-1} x}$, $\quad(\delta)$ $e^{\tan^{-1} x}$.

5. Show that if $g(x) = b_m x^m + b_{m+1}x^{m+1} + \cdots$ and $f(0) \neq 0$, then

$$\frac{f(x)}{g(x)} = \frac{a_0 + a_1 x + a_2 x^2 + \cdots}{b_m x^m + b_{m+1}x^{m+1} + \cdots} = \frac{c_{-m}}{x^m} + \frac{c_{-m+1}}{x^{m-1}} + \cdots + \frac{c_{-1}}{x} + c_0 + c_1 x + \cdots$$

and the development of the quotient has negative powers of x.

6. Develop to terms in x^6 the following functions:

(α) $\sin(k\sin x)$, $\quad(\beta)$ $\log\cos x$, $\quad(\gamma)$ $\sqrt{\cos x}$, $\quad(\delta)$ $(1 - k^2\sin^2 x)^{-\frac{1}{2}}$.

7. Carry the reversion of these series to terms in the fifth power:

(α) $y = \sin x = x - \frac{1}{6}x^3 + \cdots$, $\quad(\beta)$ $y = \tan^{-1} x = x - \frac{1}{3}x^3 + \cdots$,

(γ) $y = e^x = 1 + x + \frac{1}{2}x^2 + \cdots$, $\quad(\delta)$ $y = 2x + 3x^2 + 4x^3 + 5x^4 + \cdots$.

8. Find the smallest root of these series by the method of reversion:

$$(\alpha)\ \frac{1}{2}=\int_0^x e^{-x^2}dx = x-\frac{1}{3}x^3+\frac{1}{3!\,5}x^5-\frac{1}{3!\,7}x^7+\cdots.$$

$$(\beta)\ \frac{1}{4}=\int_0^x \cos x^2 dx. \qquad (\gamma)\ \frac{1}{10}=\int_0^x \frac{dx}{\sqrt{(1-x^2)(1-\frac{1}{4}x^2)}}.$$

9. By the formal method obtain the general equations for the coefficients in the developments of these functions and compute the first five that do not vanish:

$$(\alpha)\ \frac{\sin x}{e^x-1}, \qquad (\beta)\ \frac{2e^x}{e^x+1}, \qquad (\gamma)\ \frac{x^3}{1-2xe^x+e^{2x}}.$$

10. Obtain the general expressions for the following developments:

$$(\alpha)\ \coth x=\frac{1}{x}+\frac{x}{3}-\frac{x^3}{45}+\frac{2x^5}{945}-\cdots+\frac{B_{2n}(2x)^{2n}}{(2n)!\,x}-\cdots,$$

$$(\beta)\ \cot x=\frac{1}{x}-\frac{x}{3}-\frac{x^3}{45}-\frac{2x^5}{945}-\cdots+(-1)^n\frac{B_{2n}(2x)^{2n}}{(2n)!\,x}-\cdots.$$

$$(\gamma)\ \log\sin x=\log x-\frac{x^2}{6}-\frac{x^4}{180}-\frac{x^6}{2835}-\cdots+(-1)^n\frac{B_{2n}(2x)^{2n}}{2n\cdot(2n)!}-\cdots,$$

$$(\delta)\ \log\sinh x=\log x+\frac{x^2}{6}-\frac{x^4}{180}+\frac{x^6}{2835}-\cdots+\frac{B_{2n}(2x)^{2n}}{2n\cdot(2n)!}-\cdots.$$

11. The Eulerian numbers E_{2n} are the coefficients in the expansion of $\operatorname{sech} x$. Establish the defining equations and compute the first four as -1, 5, -61, 1385.

12. Write the expansions for $\sec x$ and $\log\tan(\frac{1}{4}\pi+\frac{1}{2}x)$.

13. From the identity $\dfrac{1}{e^x-1}-\dfrac{2}{e^{2x}-1}=\dfrac{1}{e^x+1}$ derive the expansions:

$$(\alpha)\ \frac{e^x}{e^x+1}=\frac{1}{2}+B_2(2^2-1)\frac{x}{2!}+B_4(2^4-1)\frac{x^3}{4!}+\cdots+B_{2n}(2^{2n}-1)\frac{x^{2n-1}}{2n!}+\cdots.$$

$$(\beta)\ \frac{1}{e^x+1}=\frac{1}{2}-B_2(2^2-1)\frac{x}{2!}-B_4(2^4-1)\frac{x^3}{4!}-\cdots-B_{2n}(2^{2n}-1)\frac{x^{2n-1}}{2n!}+\cdots.$$

$$(\gamma)\ \tanh x=(2^2-1)2^2B_2\frac{x}{2!}+(2^4-1)2^4B_4\frac{x^3}{4!}+\cdots+(2^{2n}-1)2^{2n}B_{2n}\frac{x^{2n-1}}{2n!}+\cdots.$$

$$(\delta)\ \tan x=x+\frac{x^3}{3}+\frac{2x^5}{15}+\frac{17x^7}{315}+\cdots+(-1)^{n-1}(2^{2n}-1)2^{2n}B_{2n}\frac{x^{2n-1}}{2n!}+\cdots.$$

$$(\epsilon)\ \log\cos x=-\frac{x^2}{2}-\frac{x^4}{12}-\frac{x^6}{45}-\cdots-(-1)^{n-1}(2^{2n}-1)2^{2n}B_{2n}\frac{x^{2n}}{2n\cdot 2n!}-\cdots.$$

$$(\zeta)\ \log\tan x=\log x+\frac{x^2}{3}+\frac{7x^4}{60}+\cdots+(-1)^{n-1}(2^{2n-1}-1)2^{2n}B_{2n}\frac{x^{2n}}{n\cdot 2n!}+\cdots.$$

$$(\eta)\ \csc x=\frac{1}{2}\left(\cot\frac{x}{2}+\tan\frac{x}{2}\right)=\frac{1}{x}+\frac{x}{3!}+\cdots+(-1)^{n-1}2(2^{2n-1}-1)B_{2n}\frac{x^{2n}}{2n!}.$$

(θ) $\log\cosh x$. $\quad(\iota)$ $\log\tanh x$. $\quad(\kappa)$ $\operatorname{csch} x$. $\quad(\lambda)$ $\sec^2 x$.

Observe that the Bernoullian numbers afford a general development for all the trigonometric and hyperbolic functions and their logarithms with the exception of the sine and cosine (which have known developments) and the secant (which requires the Eulerian numbers). The importance of these numbers is therefore apparent.

14. The coefficients $P_1(y), P_2(y), \cdots, P_n(y)$ in the development

$$\frac{e^{yx}-1}{e^x-1} = y + P_1(y)x + P_2(y)x^2 + \cdots + P_n(y)x^n + \cdots$$

are called Bernoulli's polynomials. Show that $(n+1)!\,P_n(y) = (B+y)^{n+1} - B^{n+1}$ and thus compute the first six polynomials in y.

15. If $y = N$ is a positive integer, the quotient in Ex. 14 is simple. Hence

$$n!\,P_n(N) = 1 + 2^n + 3^n + \cdots + (N-1)^n$$

is easily shown. With the aid of the polynomials found above compute:

(α) $1 + 2^4 + 3^4 + \cdots + 10^4$. $\qquad (\beta)$ $1 + 2^5 + 3^5 + \cdots + 9^5$.

(γ) $1 + 2^2 + 3^2 + \cdots + (N-1)^2$. $\qquad (\delta)$ $1 + 2^3 + 3^3 + \cdots + (N-1)^3$.

16. Interpret $\dfrac{1}{1-ax}\,\dfrac{1}{1-bx} = \dfrac{1}{x(a-b)}\left[\dfrac{1}{1-ax} - \dfrac{1}{1-bx}\right] = \sum \dfrac{a^{n+1}-b^{n+1}}{a-b}\,x^n$.

17. From $\displaystyle\int_0^\infty e^{-(1-ax)t}dt = \frac{1}{1-ax}$ establish formally

$$1 + a_1x + a_2x^2 + a_3x^3 + \cdots = \int_0^\infty e^{-t}F(xt)\,dt = \frac{1}{x}\int_0^\infty e^{-\frac{u}{x}}F(u)\,du.$$

where $\qquad F(u) = 1 + a_1u + \dfrac{1}{2!}a_2u^2 + \dfrac{1}{3!}a_3u^3 + \cdots$.

Show that the integral will converge when $0 < x < 1$ provided $|a_i| \leqq 1$.

18. If in a series the coefficients $a_i = \displaystyle\int_0^1 t^i f(t)\,dt$, show

$$1 + a_1x + a_2x^2 + a_3x^3 + \cdots = \int_0^1 \frac{f(t)}{1-xt}\,dt.$$

19. Note that Exs. 17 and 18 convert a series into an integral. Show

(α) $1 + \dfrac{x}{2^p} + \dfrac{x^2}{3^p} + \dfrac{x^3}{4^p} + \cdots = \dfrac{1}{\Gamma(p)}\displaystyle\int_0^1 \frac{(-\log t)^{p-1}}{1-xt}\,dt$ by $\dfrac{\Gamma(p)}{n^p} = \displaystyle\int_0^\infty e^{-n\xi}\xi^{p-1}d\xi$.

(β) $\dfrac{1}{1+1^2} + \dfrac{x}{1+2^2} + \dfrac{x^2}{1+3^2} + \cdots = -\displaystyle\int_0^1 \frac{\sin\log t}{1-xt}\,dt$ by $\dfrac{1}{1+n^2} = \displaystyle\int_0^\infty e^{-n\xi}\sin\xi\,d\xi$.

(γ) $1 + \dfrac{a}{b}x + \dfrac{a(a+1)}{b(b+1)}x^2 + \dfrac{a(a+1)(a+2)}{b(b+1)(b+2)}x^3 + \cdots$

$$= \frac{\Gamma(b)}{\Gamma(a)\,\Gamma(b-a)}\int_0^1 \frac{t^{a-1}(1-t)^{b-a-1}}{1-xt}\,dt.$$

20. In case the coefficients in a series are alternately positive and negative show that Euler's transformed series may be written

$$a_1x - a_2x^2 + a_3x^3 - a_4x^4 + \cdots = a_1y + \Delta a_1y^2 + \Delta^2 a_1y^3 + \Delta^3 a_1y^4 + \cdots$$

where $\Delta a_1 = a_1 - a_2$, $\Delta^2 a_1 = \Delta a_1 - \Delta a_2 = a_1 - 2a_2 + a_3, \cdots$ are the successive first, second, $\cdots$ differences of the numerical coefficients.

21. Compute the values of these series by the method of Ex. 20 with $x = 1$, $y = \frac{1}{2}$. Add the first few terms and apply the method of differences to the next few as indicated:

(α) $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = 0.69315$, add 8 terms and take 7 more,

(β) $1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \cdots = 0.6049$, add 5 terms and take 7 more.

(γ) $\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = 0.78539813$, add 10 and take 11 more.

(δ) Prove $\left(1 + \frac{1}{2^p} + \frac{1}{3^p} + \frac{1}{4^p} + \cdots\right) = \frac{2^{p-1}}{2^{p-1} - 1}\left(1 - \frac{1}{2^p} + \frac{1}{3^p} - \frac{1}{4^p} + \cdots\right)$

and compute for $p = 1.01$ with the aid of five-place tables.

22. If an infinite series converges absolutely, show that any infinite series the terms of which are selected from the terms of the given series must also converge. What if the given series converged, but not absolutely?

23. Note that the proof concerning term-by-term integration (p. 432) would not hold if the interval were infinite. Discuss this case with especial references to justifying if possible the formal evaluations of Exs. 12 (α), (δ), p. 439.

24. Check the formula of Ex. 17 by termwise integration. Evaluate

$$\frac{1}{x}\int_0^\infty e^{-\frac{u}{x}} J_0(bu)\,du - 1 - \tfrac{1}{2}b^2x^2 + \tfrac{1}{2}\cdot\tfrac{3}{2}\,\frac{b^4x^4}{2\,!} - \cdots = (1 + b^2x^2)^{-\frac{1}{2}}$$

by the inverse transformation. See Exs. 8 and 15, p. 399.

CHAPTER XVII

SPECIAL INFINITE DEVELOPMENTS

171. The trigonometric functions. If m is an odd integer, say $m = 2n + 1$, De Moivre's Theorem (§ 72) gives

$$\frac{\sin m\phi}{m \sin \phi} = \cos^{2n}\phi - \frac{(m-1)(m-2)}{3!}\cos^{2n-2}\phi \sin^2\phi + \cdots, \qquad (1)$$

where by virtue of the relation $\cos^2\phi = 1 - \sin^2\phi$ the right-hand member is a polynomial of degree n in $\sin^2\phi$. From the left-hand side it is seen that the value of the polynomial is 1 when $\sin\phi = 0$ and that the n roots of the polynomials are

$$\sin^2 \pi/m, \qquad \sin^2 2\pi/m, \qquad \cdots, \qquad \sin^2 n\pi/m.$$

Hence the polynomial may be factored in the form

$$\frac{\sin m\phi}{m \sin \phi} = \left(1 - \frac{\sin^2\phi}{\sin^2 \pi/m}\right)\left(1 - \frac{\sin^2\phi}{\sin^2 2\pi/m}\right)\cdots\left(1 - \frac{\sin^2\phi}{\sin^2 n\pi/m}\right). \quad (2)$$

If the substitutions $\phi = x/m$ and $\phi = ix/m$ be made,

$$\frac{\sin x}{m \sin x/m} = \left(1 - \frac{\sin^2 x/m}{\sin^2 \pi/m}\right)\left(1 - \frac{\sin^2 x/m}{\sin^2 2\pi/m}\right)\cdots\left(1 - \frac{\sin^2 x/m}{\sin^2 n\pi/m}\right), \quad (3)$$

$$\frac{\sinh x}{m \sinh x/m} = \left(1 + \frac{\sinh^2 x/m}{\sin^2 \pi/m}\right)\left(1 + \frac{\sinh^2 x/m}{\sin^2 2\pi/m}\right)\cdots\left(1 + \frac{\sinh^2 x/m}{\sin^2 n\pi/m}\right). \quad (3')$$

Now if m be allowed to become infinite, passing through successive odd integers, these equations remain true and it would appear that the limiting relations would hold:

$$\frac{\sin x}{x} = \left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{2^2\pi^2}\right)\cdots = \prod_1^\infty \left(1 - \frac{x^2}{k^2\pi^2}\right), \qquad (4)$$

$$\frac{\sinh x}{x} = \left(1 + \frac{x^2}{\pi^2}\right)\left(1 + \frac{x^2}{2^2\pi^2}\right)\cdots = \prod_1^\infty \left(1 + \frac{x^2}{k^2\pi^2}\right), \qquad (4')$$

since $$\lim_{m=\infty} \frac{\sin^2 \dfrac{x}{m}}{\sin^2 \dfrac{k\pi}{m}} = \lim_{m=\infty} \frac{\left(\dfrac{x}{m} - \dfrac{1}{6}\dfrac{x^3}{m^3} + \cdots\right)^2}{\left(\dfrac{k\pi}{m} - \dfrac{1}{6}\left(\dfrac{k\pi}{m}\right)^3 + \cdots\right)^2} = \frac{x^2}{k^2\pi^2}.$$

In this way *the expansions into infinite products*

$$\sin x = x \prod_1^\infty \left(1 - \frac{x^2}{k^2\pi^2}\right); \qquad \sinh x = x \prod_1^\infty \left(1 + \frac{x^2}{k^2\pi^2}\right) \tag{5}$$

would be found. As the theorem that the limit of a product is the product of the limits holds in general only for finite products, the process here followed must be justified in detail.

For the justification the consideration of $\sinh x$, which involves only positive quantities, is simpler. Take the logarithm and split the sum into two parts

$$\log \frac{\sinh x}{m \sinh \frac{x}{m}} = \sum_1^p \log \left(1 + \frac{\sinh^2 \frac{x}{m}}{\sin^2 \frac{k\pi}{m}}\right) + \sum_{p+1}^n \log \left(1 + \frac{\sinh^2 \frac{x}{m}}{\sin^2 \frac{k\pi}{m}}\right).$$

As $\log(1 + \alpha) < \alpha$, the second sum may be further transformed to

$$R = \sum_{p+1}^n \log \left(1 + \frac{\sinh^2 \frac{x}{m}}{\sin^2 \frac{k\pi}{m}}\right) < \sum_{p+1}^n \frac{\sinh^2 \frac{x}{m}}{\sin^2 \frac{k\pi}{m}} = \sinh^2 \frac{x}{m} \sum_{p+1}^n \frac{1}{\sin^2 \frac{k\pi}{m}}.$$

Now as $n < \frac{1}{2} m$, the angle $k\pi/m$ is less than $\frac{1}{2}\pi$, and $\sin \xi > 2\xi/\pi$ for $\xi < \frac{1}{2}\pi$, by Ex. 28, p. 11. Hence

$$R < \sinh^2 \frac{x}{m} \sum_{p+1}^n \frac{m^2}{4k^2} = \frac{m^2}{4} \sinh^2 \frac{x}{m} \sum_{p+1}^n \frac{1}{k^2} < \frac{m^2}{4} \sinh^2 \frac{x}{m} \int_p^\infty \frac{dk}{k^2}.$$

Hence
$$\log \frac{\sinh x}{m \sinh \frac{x}{m}} - \sum_1^p \left(1 + \frac{\sinh^2 \frac{x}{m}}{\sin^2 \frac{k\pi}{m}}\right) < \frac{m^2}{4p} \sinh^2 \frac{x}{m}.$$

Now let m become infinite. As the sum on the left is a finite, the limit is simply

$$\log \frac{\sinh x}{x} - \sum_1^p \left(1 + \frac{x^2}{k^2\pi^2}\right) < \frac{x^2}{4p}; \text{ and } \log \frac{\sinh x}{x} = \sum_1^\infty \left(1 + \frac{x^2}{k^2\pi^2}\right)$$

then follows easily by letting p become infinite. Hence the justification of (4′).

By the differentiation of the series of logarithms of (5),

$$\log \frac{\sin x}{x} = \sum_1^\infty \log \left(1 - \frac{x^2}{k^2\pi^2}\right), \qquad \log \frac{\sinh x}{x} = \sum_1^\infty \log \left(1 + \frac{x^2}{k^2\pi^2}\right), \tag{6}$$

the expressions of $\cot x$ and $\coth x$ in series of fractions

$$\cot x = \frac{1}{x} - \sum_1^\infty \frac{2x}{k^2\pi^2 - x^2}, \qquad \coth x = \frac{1}{x} + \sum_1^\infty \frac{2x}{k^2\pi^2 + x^2} \tag{7}$$

are found. And the differentiation is legitimate if these series converge uniformly. For the hyperbolic function the uniformity of the convergence follows from the M-test

$$\frac{1}{k^2\pi^2 + x^2} < \frac{1}{k^2\pi^2}, \quad \text{and} \quad \sum \frac{1}{k^2\pi^2} \text{ converges.}$$

The accuracy of the series for $\cot x$ may then be inferred by the substitution of ix for x instead of by direct examination. As

$$\frac{-2x}{k^2\pi^2 - x^2} = \frac{1}{x - k\pi} + \frac{1}{x + k\pi}, \qquad \cot x = \sum_{-\infty}^{+\infty} \frac{1}{x - k\pi}. \tag{8}$$

In this expansion, however, it is necessary still to associate the terms for $k = +n$ and $k = -n$: for each of the series for $k > 0$ and for $k < 0$ diverges.

172. In the series for $\coth x$ replace x by $\frac{1}{2}x$. Then, by (22), p. 447,

$$\frac{x}{2}\coth\frac{x}{2} = 1 + \sum_1^\infty \frac{2x^2}{4k^2\pi^2 + x^2} = 1 + \sum_1^\infty B_{2n}\frac{x^{2n}}{2n!}. \tag{9}$$

If the first series can be arranged according to powers of x, an expression for B_{2n} will be found. Consider the identity

$$\frac{t}{1+t} = -\sum_{p=1}^{n-1}(-t)^p - \frac{(-t)^n}{1+t} = -\sum_1^{n-1}(-t)^p - \theta(-t)^n,$$

which is derived by division and in which θ is a proper fraction if t is positive. Substitute $t = x^2/4k^2\pi^2$; then

$$\frac{x^2}{4k^2\pi^2 + x^2} = -\sum_1^{n-1}\left(-\frac{x^2}{4k^2\pi^2}\right)^p - \theta_k\left(-\frac{x^2}{4k^2\pi^2}\right)^n,$$

$$\frac{x}{2}\coth\frac{x}{2} - 1 = -2\sum_{k=1}^\infty\left[\sum_{p=1}^{n-1}\left(\frac{-x^2}{4k^2p^2}\right)^p - \theta_k\left(\frac{-x^2}{4k^2\pi^2}\right)^n\right]$$

$$= -2\sum_{p=1}^{n-1}\left[\left(\frac{-x^2}{4\pi^2}\right)^p\sum_{k=1}^\infty\frac{1}{k^{2p}}\right] - 2\theta\left(\frac{-x^2}{4\pi^2}\right)^n\sum_{k=1}^\infty\frac{1}{k^{2n}}.\text{*}$$

Let
$$\sum_1^\infty\frac{1}{k^{2p}} = 1 + \frac{1}{2^{2p}} + \frac{1}{3^{2p}} + \cdots = S_{2p}.$$

$$\frac{x}{2}\coth\frac{x}{2} - 1 = -2\sum_1^{n-1}S_{2p}\left(\frac{-x^2}{4\pi^2}\right)^p - 2\theta S_{2n}\left(\frac{-x^2}{4\pi^2}\right)^n.$$

* The θ is still a proper fraction since each θ_k is. The interchange of the order of summation is legitimate because the series would still converge if all signs were positive, since Σk^{-2p} is convergent.

As S_{2n} approaches 1 when n becomes infinite, the last term approaches 0 if $x < 2\pi$, and the identical expansions are

$$2\sum_1^\infty S_{2p}(-1)^{p-1}\frac{x^{2p}}{(2\pi)^{2p}} = \sum_1^\infty B_{2p}\frac{x^{2p}}{2p!} = \frac{x}{2}\coth\frac{x}{2} - 1. \tag{10}$$

Hence
$$B_{2p} = (-1)^{p-1}\frac{2(2p)!}{(2\pi)^{2p}}S_{2p} \tag{11}$$

and
$$\frac{x}{2}\coth\frac{x}{2} = 1 + \sum_1^{n-1} B_{2p}\frac{x^{2p}}{2p!} + \theta B_{2n}\frac{x^{2n}}{2n!}. \tag{12}$$

The desired expression for B_{2n} is thus found, and it is further seen that the expansion for $\frac{1}{2}x\coth\frac{1}{2}x$ can be broken off at any term with an error less than the first term omitted. This did not appear from the formal work of § 170. Further it may be noted that for large values of n the numbers B_{2n} are very large.

It was seen in treating the Γ-function that (Ex. 17, p. 385)

$$\log\Gamma(n) = (n - \tfrac{1}{2})\log n - n + \log\sqrt{2\pi} + \omega(n),$$

where
$$\omega(n) = \int_{-\infty}^0 \left(\frac{x}{2}\coth\frac{x}{2} - 1\right)e^{nx}\frac{dx}{x^2}.$$

As
$$\int_{-\infty}^0 x^{2p}e^{nx}dx = \int_0^\infty x^{2p}e^{-nx}dx = \frac{\Gamma(2p+1)}{n^{2p+1}} = \frac{2p!}{n^{2p+1}},$$

the substitution of (12), and the integration gives the result

$$\omega(n) = \frac{B_2 n^{-1}}{1\cdot 2} + \frac{B_4 n^{-3}}{3\cdot 4} + \cdots + \frac{B_{2p-2}n^{-2p+3}}{(2p-3)(2p-2)} + \frac{\theta B_{2p}n^{-2p+1}}{(2p-1)2p}. \tag{13}$$

For large values of n this development starts to converge very rapidly, and by taking a few terms a very good value of $\omega(n)$ can be obtained; but too many terms must not be taken. Compare §§ 151, 154.

EXERCISES

1. Prove $\cos x = \dfrac{\sin 2x}{2\sin x} = \prod_0^\infty\left(1 - \dfrac{4x^2}{(2k+1)^2\pi^2}\right)$.

2. On the assumption that the product for $\sinh x$ may be multiplied out and collected according to powers of x, show that

$$(\alpha)\ \sum_{k=1}^\infty\frac{1}{k^2} = \frac{\pi^2}{6}, \qquad (\beta)\ \sum_{k=1}^\infty\sum_{l=1}^\infty\frac{1}{k^2l^2} = \frac{\pi^4}{120},\ \text{where } k \neq l.$$

$$(\gamma)\ \sum_{k=1}^\infty\frac{1}{k^4} = \frac{\pi^4}{90}, \qquad (\delta)\ \sum_{k=1}^\infty\sum_{l=1}^\infty\frac{1}{k^2l^2} = \frac{\pi^4}{36},\ \text{if } k \text{ may equal } l.$$

3. By aid of Ex. 21 (δ), p. 452, show: $(\alpha)\ 1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots=\frac{\pi^2}{6}$,

$$(\beta)\ 1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\cdots=\frac{\pi^2}{8},\qquad (\gamma)\ 1-\frac{1}{2^2}+\frac{1}{3^2}-\frac{1}{4^2}+\cdots=\frac{\pi^2}{12}.$$

4. Prove: $(\alpha)\ \int_0^1 \frac{\log x}{1-x}dx=-\frac{\pi^2}{6}$, $\qquad (\beta)\ \int_0^1 \frac{\log x}{1+x}dx=-\frac{\pi^2}{12}$,

$$(\gamma)\ \int_0^1 \frac{\log x}{1-x^2}dx=-\frac{\pi^2}{8},\qquad (\delta)\ \int_0^1 \log\frac{1+x}{1-x}\frac{dx}{x}=\frac{\pi^2}{4}.$$

5. From $\qquad \tan x=-\cot\left(x-\frac{1}{2}\pi\right)=-\sum_{-\infty}^{+\infty}\frac{1}{x-(k+\frac{1}{2})\pi}$

show $\csc x=\frac{1}{2}\left(\cot\frac{x}{2}+\tan\frac{x}{2}\right)=\sum_{-\infty}^{+\infty}\frac{(-1)^k}{x-k\pi}=\frac{1}{x}+\sum_1^{\infty}\frac{(-1)^k 2x}{x^2-k^2\pi^2}$.

6. From $\qquad \frac{1}{1+x}=\sum_0^{n-1}(-x)^k+(-1)^n\frac{x^n}{1+x}=\sum_0^{n-1}(-x)^k+(-1)^n\theta x^n$

show $\int_0^1 \frac{x^{a-1}}{1+x}dx=\sum_0^{\infty}\frac{(-1)^k}{a+k}$, and compute for $a=\frac{1}{4}$ by Ex. 21, p. 452.

7. If a is a proper fraction so that $1-a$ is a proper fraction, show

$$(\alpha)\ \int_0^1 \frac{x^{-a}dx}{1+x}=\sum_1^{\infty}\frac{(-1)^k}{a-k}=\int_1^{\infty}\frac{x^{a-1}}{1+x}dx,\qquad (\beta)\ \int_0^{\infty}\frac{x^{a-1}}{1+x}dx=\frac{\pi}{\sin a\pi}.$$

8. When n is large $B_{2n}=(-1)^{n-1}4\sqrt{\pi n}\left(\frac{n}{\pi e}\right)^{2n}$ approximately (Ex. 13).

9. Expand the terms of $\frac{x}{2}\coth\frac{x}{2}=1+\sum_1^{\infty}\frac{2x^2}{4k^2\pi^2+x^2}$ by division when $x<2\pi$ and rearrange according to powers of x. Is it easy to justify this derivation of (11)?

10. Find $\omega'(n)$ by differentiating under the sign and substituting. Hence get

$$\frac{\Gamma'(n)}{\Gamma(n)}=\log n-\frac{1}{2n}-\frac{B_2}{2n^2}-\frac{B_4}{4n^4}-\cdots-\frac{B_{2p-2}}{(2p-2)n^{2p-2}}-\frac{\theta B_{2p}}{2pn^{2p}}.$$

11. From $\frac{\Gamma'(n)}{\Gamma(n)}+\gamma=\int_0^1\frac{1-\alpha^{n-1}}{1-\alpha}d\alpha$ of § 149 show that, if n is integral,

$$\frac{\Gamma'(n)}{\Gamma(n)}+\gamma=1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n-1},\quad\text{and}\quad \gamma=-\frac{\Gamma'(1)}{\Gamma(1)}=0.5772156649\cdots$$

by taking $n=10$ and using the necessary number of terms of Ex. 10.

12. Prove $\log\Gamma(n+\frac{1}{2})=n(\log n-1)+\log\sqrt{2\pi}+\omega_1(n)$, where

$$\omega_1(n)=\int_{-\infty}^0\left(\frac{1}{x}-\frac{e^{\frac{x}{2}}}{e^x-1}\right)e^{nx}\frac{dx}{x},\qquad \omega_1(n)=\omega(n)-\omega(2n),$$

$$\omega_1(n)=\frac{B_2n^{-1}}{1\cdot 2}\left(1-\frac{1}{2}\right)+\frac{B_4n^{-3}}{3\cdot 4}\left(1-\frac{1}{2^3}\right)+\frac{B_6n^{-5}}{5\cdot 6}\left(1-\frac{1}{2^5}\right)+\cdots.$$

13. Show $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{\theta}{12n}}$ or $\sqrt{2\pi}\left(\frac{n+\frac{1}{2}}{e}\right)^{n+\frac{1}{2}} e^{-\frac{\theta}{24n+12}}$. Note that the results of § 149 are now obtained rigorously.

14. From $\frac{1}{1-e^{-x}} = \sum_0^{n-1} e^{-kx} + \frac{e^{-nx}}{1-e^{-x}} = \sum_0^{n-1} e^{-kx} + \theta\frac{e^{-(n-1)x}}{x}$, and the formulas of § 149, prove the expansions

(α) $\frac{d^2}{dn^2}\log\Gamma(n) = \sum_0^{\infty}\frac{1}{(n+k)^2}$, ($\beta$) $\frac{d}{dn}\log\Gamma(n) + \gamma = \sum_0^{\infty}\left(\frac{1}{1+k} - \frac{1}{n+k}\right)$,

(γ) $\log\Gamma(n+1) + \gamma n = \sum_1^{\infty}\left(\frac{n}{k} - \log\frac{n+k}{k}\right)$, ($\delta$) $\frac{1}{\Gamma(n+1)} = e^{\gamma n}\prod_1^{\infty}\left(1+\frac{n}{k}\right)e^{-\frac{n}{k}}$.

173. Trigonometric or Fourier series. If the series

$$\begin{aligned} f(x) &= \tfrac{1}{2}a_0 + \sum_1^{\infty}(a_k\cos kx + b_k \sin kx) \\ &= \tfrac{1}{2}a_0 + a_1\cos x + a_2\cos 2x + a_3\cos 3x + \cdots \\ &\qquad + b_1\sin x + b_2 \sin 2x + b_3\sin 3x + \cdots \end{aligned} \tag{14}$$

converges over an interval of length 2π in x, say $0 \leqq x < 2\pi$ or $-\pi < x \leqq \pi$, the series will converge for all values of x and will define a periodic function $f(x+2\pi) = f(x)$ of period 2π. As

$$\int_0^{2\pi}\cos kx \sin lx dx = 0 \quad\text{and}\quad \int_0^{2\pi}\begin{matrix}\cos kx\cos lx\\ \sin kx \sin lx\end{matrix}dx = 0 \text{ or } \pi \tag{15}$$

according as $k \neq l$ or $k = l$, the coefficients in (14) may be determined formally by multiplying $f(x)$ and the series by

$$1 = \cos 0x, \qquad \cos x, \qquad \sin x, \qquad \cos 2x, \qquad \sin 2x, \cdots$$

successively and integrating from 0 to 2π. By virtue of (15) each of the integrals vanishes except one, and from that one

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx dx, \qquad b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx dx. \tag{16}$$

Conversely if $f(x)$ be a function which is defined in an interval of length 2π, and which is continuous except at a finite number of points in the interval, the numbers a_k and b_k may be computed according to (16) and the series (14) may then be constructed. If this series converges to the value of $f(x)$, there has been found an expansion of $f(x)$ over the interval from 0 to 2π in a *trigonometric or Fourier series.** The question of whether the series thus found does really converge to

* By special devices some Fourier expansions were found in Ex. 10, p. 439.

the value of the function, and whether that series can be integrated or differentiated term by term to find the integral or derivative of the function will be left for special investigation. At present it will be assumed that the function may be represented by the series, that the series may be integrated, and that it may be differentiated if the differentiated series converges.

For example let e^x be developed in the interval from 0 to 2π. Here

$$a_k = \frac{1}{\pi}\int_0^{2\pi} e^x \cos kx dx = \frac{1}{k\pi}\int_0^{2\pi k} e^{\frac{y}{k}} \cos y dy = \left[\frac{e^{\frac{y}{k}}}{\pi}\left(\frac{k\sin y + \cos y}{k^2+1}\right)\right]_0^{2\pi k}$$

or
$$a_0 = \frac{1}{\pi}e^{2\pi} - \frac{1}{\pi}, \qquad a_k = \frac{1}{\pi}e^{2\pi}\frac{1}{k^2+1} - \frac{1}{\pi}\frac{1}{k^2+1},$$

and
$$b_k = \frac{1}{\pi}\int_0^{2\pi} e^x \sin kx dx = -\frac{1}{\pi}e^{2\pi}\frac{k}{k^2+1} + \frac{1}{\pi}\frac{k}{k^2+1}.$$

Hence
$$\frac{\pi e^x}{e^{2\pi}-1} = \frac{1}{2} + \frac{1}{1^2+1}\cos x + \frac{1}{2^2+1}\cos 2x + \frac{1}{3^2+1}\cos 3x + \cdots$$
$$- \frac{1}{1^2+1}\sin x - \frac{2}{2^2+1}\sin 2x - \frac{3}{3^2+1}\sin 3x + \cdots.$$

This expansion is valid only in the interval from 0 to 2π; outside that interval the series automatically repeats that portion of the function which lies in the interval. It may be remarked that the expansion does not hold for 0 or 2π but gives the point midway in the break. Note further that if the series were differentiated the coefficient of the cosine terms would be $1 + 1/k^2$ and would not approach 0 when k became infinite, so that the series would apparently oscillate. Integration from 0 to x would give

$$\frac{\pi(e^x - 1)}{e^{2\pi}-1} = \frac{1}{2}x + \frac{1}{1^2+1}\sin x + \frac{1}{2^2+1}\frac{\sin 2x}{2} + \frac{1}{3^2+1}\frac{\sin 3x}{3} + \cdots$$
$$+ \frac{1}{1^2+1}\cos x + \frac{1}{2^2+1}\cos 2x + \frac{1}{3^2+1}\cos 3x + \cdots,$$

and the term $\frac{1}{2}x$ may be replaced by its Fourier series if desired.

As the relations (15) hold not only when the integration is from 0 to 2π but also when it is over any interval of 2π from α to $\alpha + 2\pi$, the function may be expanded into series in the interval from α to $\alpha + 2\pi$ by using these values instead of 0 and 2π as limits in the formulas (16) for the coefficients. It may be shown that a function may be expanded in only one way into a trigonometric series (14) valid for an interval of length 2π; but the proof is somewhat intricate and will not be given here. If, however, the expansion of the function is desired for an interval $\alpha < x < \beta$ less than 2π, there are an infinite number of developments (14) which will answer; for if $\phi(x)$ be a

function which coincides with $f(x)$ during the interval $\alpha < x < \beta$, over which the expansion of $f(x)$ is desired, and which has any value whatsoever over the remainder of the interval $\beta < x < \alpha + 2\pi$, the expansion of $\phi(x)$ from α to $\alpha + 2\pi$ will converge to $f(x)$ over the interval $\alpha < x < \beta$.

In practice it is frequently desirable to restrict the interval over which $f(x)$ is expanded to a length π, say from 0 to π, and to seek an expansion in terms of sines or cosines alone. Thus suppose that in the interval $0 < x < \pi$ the function $\phi(x)$ be identical with $f(x)$, and that in the interval $-\pi < x < 0$ it be equal to $f(-x)$; that is, the function $\phi(x)$ is an even function, $\phi(x) = \phi(-x)$, which is equal to $f(x)$ in the interval from 0 to π. Then

$$\int_{-\pi}^{+\pi} \phi(x) \cos kx dx = 2\int_0^{\pi} \phi(x) \cos kx dx = 2\int_0^{\pi} f(x) \cos kx dx,$$

$$\int_{-\pi}^{+\pi} \phi(x) \sin kx dx = \int_0^{\pi} \phi(x) \sin kx dx - \int_0^{\pi} \phi(x) \sin kx dx = 0.$$

Hence for the expansion of $\phi(x)$ from $-\pi$ to $+\pi$ the coefficients b_k all vanish and the expansion is in terms of cosines alone. As $f(x)$ coincides with $\phi(x)$ from 0 to π, the expansion

$$f(x) = \sum_0^{\infty} a_k \cos kx, \qquad a_k = \frac{2}{\pi}\int_0^{\pi} f(x) \cos kx dx \tag{17}$$

of $f(x)$ in terms of cosines alone, and valid over the interval from 0 to π, has been found. In like manner the expansion

$$f(x) = \sum_1^{\infty} b_k \sin kx, \qquad b_k = \frac{2}{\pi}\int_0^{\pi} f(x) \sin kx dx \tag{18}$$

in term of sines alone may be found by taking $\phi(x)$ equal to $f(x)$ from 0 to π and equal to $-f(-x)$ from 0 to $-\pi$.

Let $\frac{1}{2}x$ be developed into a series of sines and into a series of cosines valid over the interval from 0 to π. For the series of sines

$$b_k = \frac{2}{\pi}\int_0^{\pi} \frac{1}{2} x \sin kx dx = -\frac{(-1)^k}{k}, \qquad \frac{x}{2} = \sum_1^{\infty} \mp \frac{\sin kx}{k}$$

or

$$\tfrac{1}{2}x = \sin x - \tfrac{1}{2}\sin 2x + \tfrac{1}{3}\sin 3x - \tfrac{1}{4}\sin 4x + \cdots. \tag{A}$$

Also $$a_0 = \frac{2}{\pi}\int_0^{\pi} \frac{1}{2} x dx = \frac{\pi}{2}, \qquad a_k = \frac{2}{\pi}\int_0^{\pi} \frac{1}{2} x \cos kx dx = \begin{cases} 0, & k \text{ even} \\ -\dfrac{2}{\pi k}, & k \text{ odd}. \end{cases}$$

Hence $$\frac{1}{2}x = \frac{\pi}{4} - \frac{2}{\pi}\left[\cos x + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \frac{\cos 7x}{7^2} + \cdots\right]. \tag{B}$$

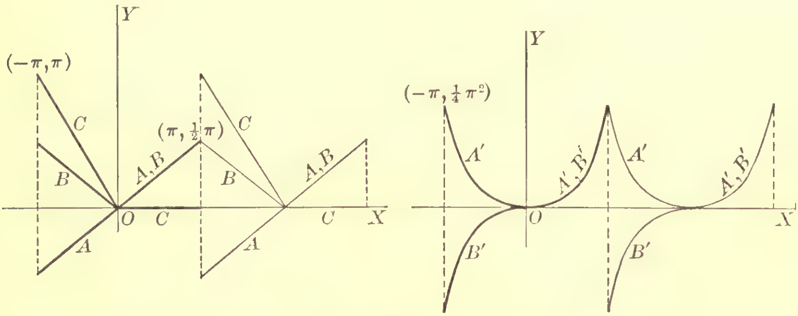
Although the two expansions define the same function $\frac{1}{2}x$ over the interval 0 to π, they will define different functions in the interval 0 to $-\pi$, as in the figure.

The development for $\frac{1}{4}x^2$ may be had by integrating either series (A) or (B).

$$\frac{1}{4}x^2 = 1 - \cos x - \frac{1}{4}(1 - \cos 2x) + \frac{1}{9}(1 - \cos 3x) - \frac{1}{16}(1 - \cos 4x) + \cdots$$

$$= \frac{\pi}{4}x - \frac{2}{\pi}\left[\sin x + \frac{\sin 3x}{3^3} + \frac{\cos 5x}{5^3} + \cdots\right].$$

These are not yet Fourier series because of the terms $\frac{1}{4}\pi x$ and the various 1's. For $\frac{1}{4}\pi x$ its sine series may be substituted and the terms $1 - \frac{1}{4} + \frac{1}{9} - \cdots$ may be collected by Ex. 3, p. 457. Hence



$$\frac{1}{4}x^2 = \frac{\pi^2}{12} - \cos x + \frac{1}{4}\cos 2x - \frac{1}{9}\cos 3x + \frac{1}{16}\cos 4x - \cdots \qquad (A')$$

or $$\frac{1}{4}x^2 = \frac{2}{\pi}\left[\left(\frac{\pi^2}{4} - 1\right)\sin x - \frac{\pi^2}{2}\sin 2x + \left(\frac{\pi^2}{12} - \frac{1}{3^2}\right)\sin 3x - \frac{\pi^2}{4}\sin 4x + \cdots\right]. \qquad (B')$$

The differentiation of the series (A) of sines will give a series in which the individual terms do not approach 0; the differentiation of the series (B) of cosines gives

$$\frac{1}{4}\pi = \sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \frac{1}{7}\sin 7x + \cdots$$

and that this is the series for $\pi/4$ may be verified by direct calculation. The difference of the two series (A) and (B) is a Fourier series

$$f(x) = \frac{\pi}{4} - \frac{2}{\pi}\left[\cos x + \frac{\cos 3x}{3^2} + \cdots\right] - \left[\sin x - \frac{\sin 2x}{2} + \cdots\right] \qquad (C)$$

which defines a function that vanishes when $0 < x < \pi$ but is equal to $-x$ when $0 > x > -\pi$.

174. For discussing the convergence of the trigonometric series as formally calculated, the sum of the first $2n+1$ terms may be written as

$$S_n = \frac{1}{\pi}\int_0^{2\pi}\left[\frac{1}{2} + \cos(t-x) + \cos 2(t-x) + \cdots + \cos n(t-x)\right]f(t)\,dt$$

$$= \frac{1}{\pi}\int_0^{2\pi}\frac{\sin(2n+1)\dfrac{t-x}{2}}{2\sin\dfrac{t-x}{2}}f(t)\,dt = \frac{1}{\pi}\int_{-\frac{x}{2}}^{\pi-\frac{x}{2}} f(x+2u)\frac{\sin(2n+1)u}{\sin u}\,du,$$

where the first step was to combine $a_k \cos kx$ and $b_k \sin kx$ after replacing x in the definite integrals (16) by t to avoid confusion, then summing by the formula of Ex. 9, p. 30, and finally changing the variable to $u = \frac{1}{2}(t - x)$. The sum S_n is therefore represented as a definite integral whose limit must be evaluated as n becomes infinite.

Let the restriction be imposed upon $f(x)$ that it shall be of limited variation in the interval $0 < x < 2\pi$. As the function $f(x)$ is of limited variation, it may be regarded as the difference $P(x) - N(x)$ of two positive limited functions which are constantly increasing and which will be continuous wherever $f(x)$ is continuous (§ 127). If $f(x)$ is discontinuous at $x = x_0$, it is still true that $f(x)$ approaches a limit, which will be denoted by $f(x_0 - 0)$ when x approaches x_0 from below; for each of the functions $P(x)$ and $N(x)$ is increasing and limited and hence each must approach a limit, and $f(x)$ will therefore approach the difference of the limits. In like manner $f(x)$ will approach a limit $f(x_0 + 0)$ as x approaches x_0 from above. Furthermore as $f(x)$ is of limited variation the integrals required for S_n, a_k, b_k will all exist and there will be no difficulty from that source. It will now be shown that

$$\lim_{n\to\infty} S_n(x_0) = \lim_{n\to\infty} \frac{1}{\pi}\int_{-\frac{x_0}{2}}^{\pi-\frac{x_0}{2}} f(x_0 + 2u)\frac{\sin(2n+1)u}{\sin u}\,du = \frac{1}{2}\left[f(x_0+0) - f(x_0-0)\right].$$

This will show that *the series converges to the function wherever the function is continuous and to the mid-point of the break wherever the function is discontinuous.*

$$\text{Let}\quad f(x_0 + 2u)\frac{\sin(2n+1)u}{\sin u} = f(x_0+2u)\frac{u}{\sin u}\frac{\sin(2n+1)u}{u} = F(u)\frac{\sin ku}{u},$$

$$\text{then}\quad S_n(x_0) = \frac{1}{\pi}\int_{-\frac{x_0}{2}}^{\pi-\frac{x_0}{2}} F(u)\frac{\sin ku}{u}\,du = \frac{1}{\pi}\int_a^b F(u)\frac{\sin ku}{u}\,du, \quad -\pi < a < 0 < b < \pi.$$

As $f(x)$ is of limited variation provided $-\pi < a \leqq u \leqq b < \pi$, so must $f(x_0 + 2u)$ be of limited variation and also $F(u) = uf/\sin u$. Then $F(u)$ may be regarded as the difference of two constantly increasing positive functions, or, if preferable, of two constantly decreasing positive functions; and it will be sufficient to investigate the integral of $F(u)\,u^{-1}\sin ku$ under the hypothesis that $F(u)$ is constantly decreasing. Let n be the number of times $2\pi/k$ is contained in b.

$$\int_0^b F(u)\frac{\sin ku}{u}\,du = \int_0^{\frac{2\pi}{k}} + \int_{\frac{2\pi}{k}}^{\frac{4\pi}{k}} + \cdots + \int_{\frac{2(n-1)\pi}{k}}^{\frac{2n\pi}{k}} + \int_{\frac{2n\pi}{k}}^{b} F(u)\frac{\sin ku}{u}\,du$$

$$= \int_0^{2\pi} + \int_{2\pi}^{4\pi} + \cdots + \int_{2(n-1)\pi}^{2n\pi} F\left(\frac{u}{k}\right)\frac{\sin u}{u}\,du + \int_{\frac{2n\pi}{k}}^{b} F(u)\frac{\sin ku}{u}\,du.$$

As $F(u)$ is a decreasing function, so is $u^{-1}F(u/k)$, and hence each of the integrals which extends over a complete period 2π will be positive because the negative elements are smaller than the corresponding positive elements. The integral from $2n\pi/k$ to b approaches zero as k becomes infinite. Hence for large values of k,

$$\int_0^b F(u)\frac{\sin ku}{u}\,du > \int_0^{2p\pi} F\left(\frac{u}{k}\right)\frac{\sin u}{u}\,du, \qquad p \text{ fixed and less than } n.$$

Again, $\int_0^b F(u)\frac{\sin ku}{u}du = \int_0^\pi + \int_\pi^{3\pi} + \int_{3\pi}^{5\pi}$

$$+\cdots+\int_{(2n-3)\pi}^{(2n-1)\pi} F\left(\frac{u}{k}\right)\frac{\sin u}{u}du + \int_{(2n-1)\pi}^{b} F(u)\frac{\sin ku}{u}du.$$

Here all the terms except the first and last are negative because the negative elements of the integrals are larger than the positive elements. Hence for k large,

$$\int_0^b F(u)\frac{\sin ku}{u}du < \int_0^{(2p-1)\pi} F\left(\frac{u}{k}\right)\frac{\sin u}{u}du, \qquad p \text{ fixed and less than } n.$$

In the inequalities thus established let k become infinite. Then $u/k \doteq 0$ from above and $F(u/k) \doteq F(+0)$. It therefore follows that

$$F(+0)\int_0^{(2p-1)\pi}\frac{\sin u}{u}du < \lim_{k=\infty}\int_0^b F(u)\frac{\sin ku}{u}du > F(+0)\int_0^{2p\pi}\frac{\sin u}{u}du.$$

Although p is fixed, there is no limit to the size of the number at which it is fixed. Hence the inequality may be transformed into an equality

$$\lim_{k=\infty}\int_0^b F(u)\frac{\sin ku}{u}du = F(+0)\int_0^\infty\frac{\sin u}{u}du = \frac{\pi}{2}F(+0).$$

Likewise $\quad \displaystyle\lim_{k=\infty}\int_a^0 F(u)\frac{\sin ku}{u}du = F(-0)\int_0^\infty \frac{\sin u}{u}du = \frac{\pi}{2}F(-0).$

Hence $\quad \displaystyle\lim_{k=\infty}\int_a^b F(u)\frac{\sin ku}{u}du = \frac{\pi}{2}[F(+0)+F(-0)]$

or $\quad \displaystyle\lim_{n=\infty}\frac{1}{\pi}\int_{-\frac{x_0}{2}}^{\pi-\frac{x_0}{2}} f(x_0+2u)\frac{\sin(2n+1)u}{\sin u}du = \frac{1}{2}[f(x_0+0)+f(x_0-0)].$

Hence for every point x_0 in the interval $0 < x < 2\pi$ the series converges to the function where continuous, and to the mid-point of the break where discontinuous.

As the function $f(x)$ has the period 2π, it is natural to suppose that the convergence at $x = 0$ and $x = 2\pi$ will not differ materially from that at any other value, namely, that it will be to the value $\frac{1}{2}[f(+0)+f(2\pi-0)]$. This may be shown by a transformation. If k is an odd integer, $2n+1$,

$$\sin(2n+1)u = \sin(2n+1)(\pi-u) = \sin(2n+1)u'.$$

$$\lim_{n=\infty}\int_b^\pi F(u)\frac{\sin(2n+1)u}{u}du = \lim_{n=\infty}\int_0^{\pi-b}F(u')\frac{\sin(2n+1)u'}{u'}du' = \frac{\pi}{2}F(u'=+0).$$

Hence $\quad \displaystyle\lim_{n=\infty}\int_0^\pi F(u)\frac{\sin(2n+1)u}{u}du = \lim_{n=\infty}\int_0^b + \int_b^\pi = \frac{\pi}{2}[F(+0)+F(\pi-0)].$

Now for $x = 0$ or $x = 2\pi$ the sum $S_n = \dfrac{1}{\pi}\displaystyle\int_0^\pi f(2u)\frac{\sin(2n+1)u}{\sin u}du$, and the limit will therefore be $\frac{1}{2}[f(+0)+f(2\pi-0)]$ as predicted above.

The convergence may be examined more closely. In fact

$$S_n(x) = \frac{1}{\pi}\int_{-\frac{x}{2}}^{\pi-\frac{x}{2}} f(x+2u)\frac{u}{\sin u}\frac{\sin ku}{u}du = \frac{1}{\pi}\int_{a(x)}^{b(x)}F(x,u)\frac{\sin ku}{u}du.$$

Suppose $0 < \alpha \leqq x \leqq \beta < 2\pi$ so that the least possible upper limit $b(x)$ is $\pi - \frac{1}{2}\beta$ and the greatest possible lower limit $a(x)$ is $-\frac{1}{2}\alpha$. Let n be the number of times $2\pi/k$ is contained in $\pi - \frac{1}{2}\beta$. Then for all values of x in $\alpha \leqq x \leqq \beta$,

$$\int_0^{(2p-1)\pi} F\left(x, \frac{u}{k}\right)\frac{\sin u}{u}\,du + \epsilon < \int_0^{b(x)} F(x, u)\frac{\sin ku}{u}\,du$$

$$< \int_0^{2p\pi} F\left(x, \frac{u}{k}\right)\frac{\sin u}{u}\,du + \eta, \qquad p < n,$$

where ϵ and η are the integrals over partial periods neglected above and are uniformly small for all x's of $\alpha \leqq x \leqq \beta$ since $F(x, u)$ is everywhere finite. This shows that the number p may be chosen uniformly for all x's in the interval and yet ultimately may be allowed to become infinite. If it be now assumed that $f(x)$ is continuous for $\alpha \leqq x \leqq \beta$, then $F(x, u)$ will be continuous and hence uniformly continuous in (x, u) for the region defined by $\alpha \leqq x \leqq \beta$ and $-\frac{1}{2}x \leqq u \leqq \pi - \frac{1}{2}x$. Hence $F(x, u/k)$ will converge uniformly to $F(x, +0)$ as k becomes infinite. Hence

$$F(x, +0)\int_0^{\infty}\frac{\sin u}{u}\,du + \epsilon' < \int_0^{b(x)} F(x, u)\frac{\sin ku}{u}\,du < F(x, +0)\int_0^{\infty}\frac{\sin u}{u}\,du + \eta'$$

where, if $\delta > 0$ is given, K may be taken so large that $|\epsilon'| < \delta$ and $|\eta'| < \delta$ for $k > K$; with a similar relation for the integration from $a(x)$ to 0. Hence in any interval $0 < \alpha \leqq x \leqq \beta < 2\pi$ over which $f(x)$ is continuous $S_n(x)$ converges uniformly toward its limit $f(x)$. Over such an interval the series may be integrated term by term. If $f(x)$ has a finite number of discontinuities, the series may still be integrated term by term throughout the interval $0 \leqq x \leqq 2\pi$ because $S_n(x)$ remains always finite and limited and such discontinuities may be disregarded in integration.

EXERCISES

1. Obtain the expansions over the indicated intervals. Integrate the series. Also discuss the differentiated series. Make graphs.

(α) $$\frac{\pi e^x}{2\sinh\pi} = \frac{1}{2} - \frac{1}{2}\cos x + \frac{1}{5}\cos 2x - \frac{1}{10}\cos 3x + \frac{1}{17}\cos 4x - \cdots$$
$$+ \frac{1}{2}\sin x - \frac{2}{5}\sin 2x + \frac{3}{10}\sin 3x - \frac{4}{17}\sin 4x + \cdots, \qquad -\pi \text{ to } +\pi,$$

(β) $\frac{1}{4}\pi$, as sine series, 0 to π, $\qquad (\gamma)$ $\frac{1}{4}\pi$, as cosine series, 0 to π,

(δ) $\sin x = \dfrac{4}{\pi}\left[\dfrac{1}{2} - \dfrac{\cos 2x}{1\cdot 3} - \dfrac{\cos 4x}{3\cdot 5} - \dfrac{\cos 6x}{5\cdot 7} - \cdots\right]$, 0 to π.

(ϵ) $\cos x$, as sine series, 0 to π, $\qquad (\zeta)$ e^x, as cosine series, 0 to π,

(η) $x\sin x$, $-\pi$ to π, $\qquad (\theta)$ $x\cos x$, $-\pi$ to π, $\qquad (\iota)$ $\pi + x$, $-\pi$ to π,

(κ) $\sin\theta x$, $-\pi$ to π, θ fractional, $\qquad (\lambda)$ $\cos\theta x$, $-\pi$ to π, θ fractional,

(μ) $f(x) = \begin{cases}\frac{1}{4}\pi, & 0 < x < \pi,\\ 0, & \pi < x < 2\pi,\end{cases}$ $\qquad (\nu)$ $f(x) = \begin{cases}\frac{1}{4}\pi, & 0 < x < \frac{1}{2}\pi,\\ -\frac{1}{4}\pi, & \frac{1}{2}\pi < x < \pi,\end{cases}$ as a sine series, 0 to π,

(o) $-\log\left(2\sin\dfrac{x}{2}\right) = \cos x + \dfrac{1}{2}\cos 2x + \dfrac{1}{3}\cos 3x + \dfrac{1}{4}\cos 4x + \cdots$, 0 to π.

(π) x, $-\frac{1}{2}\pi$ to $\frac{3}{2}\pi$, (ρ) $\sin\frac{1}{2}x$, $-\frac{1}{2}\pi$ to $\frac{3}{2}\pi$, (σ) $\cos\frac{1}{2}x$, $-\frac{3}{2}\pi$ to $\frac{1}{2}\pi$,

(τ) from (o) find expansions for $\log\cos\frac{1}{2}x$, $\log\operatorname{vers} x$, $\log\tan\frac{1}{2}x$. Note that in these cases, as in (o), the function does not remain finite, but its integral does.

2. What peculiarities occur in the trigonometric development from $-\pi$ to π for an odd function for which $f(x)=f(\pi-x)$? for an even function for which $f(x)=f(\pi-x)$?

3. Show that $f(x)=\sum_1^\infty b_k \sin\frac{k\pi x}{c}$ with $b_k=\frac{2}{c}\int_0^c f(x)\sin\frac{k\pi x}{c}dx$ is the trigonometric sine series for $f(x)$ over the interval $0<x<c$ and that the function thus defined is odd and of period $2c$. Write the corresponding results for the cosine series and for the general Fourier series.

4. Obtain Nos. 808–812 of Peirce's Tables. Graph the sum of Nos. 809 and 810.

5. Let $e(x)=f(x)-\frac{1}{2}a_0-a_1\cos x-\cdots-a_n\cos nx-b_1\sin x-\cdots-b_n\sin nx$ be the error made by taking for $f(x)$ the first $2n+1$ terms of a trigonometric series. The mean value of the square of $e(x)$ is $\frac{1}{2\pi}\int_{-\pi}^{+\pi}[e(x)]^2dx$ and is a function $F(a_0, a_1, \cdots, a_n, b_1, \cdots, b_n)$ of the coefficients. Show that if this mean square error is to be as small as possible, the constants $a_0, a_1, \cdots, a_n, b_1, \cdots, b_n$ must be precisely those given by (16); that is, show that (16) is equivalent to

$$\frac{\partial F}{\partial a_0}=\frac{\partial F}{\partial a_1}=\cdots=\frac{\partial F}{\partial a_n}=\frac{\partial F}{\partial b_1}=\cdots=\frac{\partial F}{\partial b_n}=0.$$

6. By using the variable λ in place of x in (16) deduce the equations

$$f(x)=\frac{1}{2\pi}\int_0^{2\pi}f(\lambda)\cos 0(\lambda-x)\,d\lambda+\frac{1}{\pi}\sum_1^\infty\int_0^{2\pi}f(\lambda)\cos k(\lambda-x)\,d\lambda$$

$$=\frac{1}{2\pi}\sum_{-\infty}^{\infty}\int_0^{2\pi}f(\lambda)e^{\pm k(\lambda-x)i}d\lambda=\frac{1}{2\pi}\sum_{-\infty}^{\infty}e^{\mp kxi}\int_0^{2\pi}f(x)e^{\pm kxi}dx;$$

and hence infer $\quad f(x)=\sum_{-\infty}^{\infty}\alpha_k e^{\mp kxi}, \qquad \alpha_k=\frac{1}{2\pi}\int_0^{2\pi}f(x)e^{\pm kxi}dx.$

7. Without attempting rigorous analysis show formally that

$$\int_{-\infty}^{\infty}\phi(\alpha)d\alpha=\lim_{\Delta\alpha\doteq 0}[\cdots+\phi(-n\cdot\Delta\alpha)\Delta\alpha+\phi(-n+1\cdot\Delta\alpha)\Delta\alpha+\cdots+\phi(-1\cdot\Delta\alpha)\Delta\alpha$$
$$+\phi(0\cdot\Delta\alpha)\Delta\alpha+\phi(1\cdot\Delta\alpha)\Delta\alpha+\cdots+\phi(n\cdot\Delta\alpha)\Delta\alpha+\cdots]$$

$$=\lim_{\Delta\alpha\doteq 0}\sum_{-\infty}^{\infty}\phi(k\cdot\Delta\alpha)\Delta\alpha=\lim_{c\doteq\infty}\sum_{-\infty}^{\infty}\phi\left(k\frac{a}{c}\right)\frac{a}{c}.$$

Show $\quad f(x)=\frac{1}{2c}\sum_{-\infty}^{\infty}\int_{-c}^{c}f(\lambda)e^{\pm\frac{k\pi}{c}(\lambda-x)i}d\lambda=\frac{1}{2\pi}\sum_{-\infty}^{\infty}\int_{-c}^{c}f(\lambda)e^{\pm\frac{k\pi}{c}(\lambda-x)i}\frac{\pi}{c}d\lambda$

is the expansion of $f(x)$ by Fourier series from $-c$ to c. Hence infer that

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\lambda)e^{\pm\alpha(\lambda-x)i}d\lambda d\alpha=\lim_{c=\infty}\frac{1}{2\pi}\sum_{-\infty}^{\infty}\int_{-c}^{c}f(\lambda)e^{\pm\frac{k\pi}{c}(\lambda-x)i}d\lambda\frac{\pi}{c}$$

is an expression for $f(x)$ as a double integral, which may be expected to hold for all values of x. Reduce this to the form of a Fourier Integral (Ex. 15, p. 377)

$$f(x) = \frac{1}{\pi}\int_0^{\infty}\int_{-\infty}^{\infty} f(\lambda)\cos\alpha(\lambda - x)\,d\lambda\,d\alpha.$$

8. Assume the possibility of expanding $f(x)$ between -1 and $+1$ as a series of Legendre polynomials (Exs. 13–20, p. 252, Ex. 16, p. 440) in the form

$$f(x) = a_0P_0(x) + a_1P_1(x) + a_2P_2(x) + \cdots + a_nP_n(x) + \cdots.$$

By the aid of Ex. 19, p. 253, determine the coefficients as $a_k = \dfrac{2k+1}{2}\displaystyle\int_{-1}^{1} f(x)\,P_k(x)\,dx.$ For this expansion, form $e(x)$ as in Ex. 5 and show that the determination of the coefficients a_i so as to give a least mean square error agrees with the determination here found.

9. Note that the expansion of Ex. 8 represents a function $f(x)$ between the limits ± 1 as a polynomial of the nth degree in x, plus a remainder. It may be shown that precisely this polynomial of degree n gives a smaller mean square error over the interval than any other polynomial of degree n. For suppose

$$g_n(x) = c_0 + c_1x + \cdots + c_nx^n = b_0 + b_1P_1 + \cdots + b_nP_n$$

be any polynomial of degree n and its equivalent expansion in terms of Legendre polynomials. Now if the c's are so determined that the mean value of $[f(x) - g_n(x)]^2$ is a minimum, so are the b's, which are linear homogeneous functions of the c's. Hence the b's must be identical with the a's above. Note that whereas the Maclaurin expansion replaces $f(x)$ by a polynomial in x which is a very good approximation near $x = 0$, the Legendre expansion replaces $f(x)$ by a polynomial which is the best expansion when the whole interval from -1 to $+1$ is considered.

10. Compute (cf. Ex. 17, p. 252) the polynomials $P_1 = x$, $P_2 = -\frac{1}{2} + \frac{3}{2}x^2$,

$$P_3 = -\tfrac{3}{2}x + \tfrac{5}{2}x^3, \qquad P_4 = \tfrac{3}{8} - \tfrac{15}{4}x^2 + \tfrac{35}{8}x^4, \qquad P_5 = \tfrac{15}{8}x - \tfrac{35}{4}x^3 + \tfrac{63}{8}x^5.$$

Compute $\displaystyle\int_{-1}^{1} x^i \sin\pi x\,dx = 0,\ \frac{2}{\pi}\left(1 - \frac{6}{\pi^2}\right),\ 0,\ \frac{2}{\pi},\ 0$ when $i = 4, 3, 2, 1, 0$. Hence show that the polynomial of the fourth degree which best represents $\sin\pi x$ from -1 to $+1$ reduces to degree three, and is

$$\sin\pi x = \frac{3}{\pi}x - \frac{7}{\pi}\left(\frac{15}{\pi^2} - 1\right)\left(\frac{5}{2}x^3 - \frac{3}{2}x\right) = 2.69x - 2.89x^3.$$

Show that the mean square error is 0.004 and compare with that due to Maclaurin's expansion if the term in x^4 is retained or if the term in x^3 is retained.

11. Expand $\sin\frac{1}{2}\pi x = \dfrac{12}{\pi^2}P_1 - \dfrac{168}{\pi^2}\left(\dfrac{10}{\pi^2} - 1\right)P_3 = 1.553x - 0.562x^3.$

12. Expand from -1 to $+1$, as far as indicated, these functions:

(α) $\cos\pi x$	to P_4,	(β) e^x	to P_5,	(γ) $\log(1 + x)$	to P_4,
(δ) $\sqrt{1 - x^2}$	to P_4,	(ϵ) $\cos^{-1}x$	to P_4,	(ζ) $\tan^{-1}x$	to P_5,
(η) $\dfrac{1}{\sqrt{1 + x}}$	to P_3,	(θ) $\dfrac{1}{\sqrt{1 - x^2}}$	to P_3,	(ι) $\dfrac{1}{\sqrt{1 + x^2}}$	to P_3.

What simplifications occur if $f(x)$ is odd or if it is even?

175. The Theta functions. It has been seen that a function with the period 2π may be expanded into a trigonometric series; that if the function is odd, the series contains only sines; and if, furthermore, the function is symmetric with respect to $x = \frac{1}{2}\pi$, the odd multiples of the angle will alone occur. In this case let

$$f(x) = 2\left[a_0 \sin x - a_1 \sin 3x + \cdots + (-1)^n a_n \sin(2n+1)x + \cdots\right].$$

As $2 \sin nx = -i(e^{nxi} - e^{-nxi})$, the series may be written

$$f(x) = 2\sum_{0}^{\infty} (-1)^n a_n \sin(2n+1)x = -i\sum_{-\infty}^{\infty}(-1)^n a_n e^{(2n+1)xi}, \quad a_{-n} = a_{n-1}.$$

This exponential form is very convenient for many purposes. Let $i\rho$ be added to x. The general term of the series is then

$$a_{n-1}e^{(2n-1)(x+i\rho)i} = a_{n-1}e^{-(2n-1)\rho}e^{-2xi}e^{(2n+1)xi}.$$

Hence if the coefficients of the series satisfy $a_{n-1}e^{-2n\rho} = a_n$, the new general term is identical with the succeeding term in the given series multiplied by $-e^{\rho}e^{-2xi}$. Hence

$$f(x + i\rho) = -e^{\rho}e^{-2xi}f(x) \quad \text{if} \quad a_{n-1} = a_n e^{2n\rho}.$$

The recurrent relation between the coefficients will determine them in terms of a_0. For let $q = e^{-\rho}$. Then

$$a_n = a_{n-1}q^{2n} = a_{n-2}q^{2n}q^{2n-2} = \cdots = a_0 q^{2n}q^{2n-2}\cdots q^2 = a_0 q^{n^2+n},$$

$$a_0 = a_{-1} = a_{-2}q^{-2} = a_{-3}q^{-2}q^{-4} = \cdots = a_{-n-1}q^{-n^2-n}.$$

The new relation on the coefficients is thus compatible with the original relation $a_{-n} = a_{n-1}$. If $a_0 = q^{\frac{1}{4}}$, the series thus becomes

$$f(x) = 2q^{\frac{1}{4}}\sin x - 2q^{\frac{9}{4}}\sin 3x + \cdots + (-1)^n 2q^{\frac{1}{4}(2n+1)^2}\sin(2n+1)x + \cdots,$$

$$f(x+2\pi) = f(x), \quad f(x+\pi) = -f(x), \quad f(x+i\rho) = -q^{-1}e^{-2xi}f(x).$$

The function thus defined formally has important properties.

In the first place it is important to discuss the convergence of the series. Apply the test ratio to the exponential form.

$$a_{n+1}/a_n = q^{2n}e^{2xi}, \qquad a_{-n-1}/a_{-n} = q^{2n}e^{-2xi}.$$

For any x this ratio will approach the limit 0 if q is numerically less than 1. Hence the series converges for all values of x provided $|q| < 1$. Moreover if $|x| < \frac{1}{2}G$, the absolute value of the ratio is less than $|q|^{2n}e^{G}$, which approaches 0 as n becomes infinite. The terms of the series therefore ultimately become less than those of any assigned geometric

series. This establishes the uniform convergence and consequently the continuity of $f(x)$ for all real or complex values of x. As the series for $f'(x)$ may be treated similarly, the function has a continuous derivative and is everywhere analytic.

By a change of variable and notation let

$$H(u) = f\left(\frac{\pi u}{2K}\right), \qquad q = e^{-\pi\frac{K'}{K}}, \tag{19}$$

$$H(u) = 2q^{\frac{1}{4}}\sin\frac{\pi u}{2K} - 2q^{\frac{9}{4}}\sin\frac{3\pi u}{2K} + 2q^{\frac{25}{4}}\sin\frac{5\pi u}{2K} - \cdots. \tag{20}$$

The function $H(u)$, called eta of u, has therefore the properties

$$H(u + 2K) = -H(u), \qquad H(u + 2iK') = -q^{-1}e^{-\frac{i\pi}{K}u}H(u), \tag{21}$$

$$H(u + 2mK + 2inK') = (-1)^{m+n}q^{-n}e^{-\frac{in\pi}{K}u}H(u), \qquad m, n \text{ integers.}$$

The quantities $2K$ and $2iK'$ are called the *periods* of the function. They are not true periods in the sense that 2π is a period of $f(x)$; for when $2K$ is added to u, the function does not return to its original value, but is changed in sign; and when $2iK'$ is added to u, the function takes the multiplier written above.

Three new functions will be formed by adding to u the quantity K or iK' or $K + iK'$, that is, the *half periods*, and making slight changes suggested by the results. First let $H_1(u) = H(u + K)$. By substitution in the series (20),

$$H_1(u) = 2q^{\frac{1}{4}}\cos\frac{\pi u}{2K} + 2q^{\frac{9}{4}}\cos\frac{3\pi u}{2K} + 2q^{\frac{25}{4}}\cos\frac{5\pi u}{2K} + \cdots. \tag{22}$$

By using the properties of H, corresponding properties of H_1,

$$H_1(u + 2K) = -H_1(u), \qquad H_1(u + 2iK') = +q^{-1}e^{-\frac{i\pi}{K}u}H_1(u), \tag{23}$$

are found. Second let iK' be added to u in $H(u)$. Then

$$q^{\frac{1}{4}(2n+1)^2}e^{(2n+1)\frac{\pi i}{2K}(u+iK')} = q^{n^2+n+\frac{1}{4}}e^{-\pi(n+\frac{1}{2})\frac{K'}{K}}e^{(2n+1)\frac{\pi i}{2K}u}$$

is the general term in the exponential development of $H(u + iK')$ apart from the coefficient $\pm i$. Hence

$$H(u + iK') = i\sum_{-\infty}^{\infty}(-1)^n q^{n^2-\frac{1}{4}}e^{-\frac{\pi i}{2K}u}e^{2n\frac{\pi i}{2K}u}$$

$$= iq^{-\frac{1}{4}}e^{-\frac{\pi i}{2K}u}\sum_{-\infty}^{\infty}(-1)^n q^{n^2}e^{2n\frac{\pi i}{2K}u}$$

Let $$\Theta(u) = -iq^{\frac{1}{4}}e^{\frac{i\pi}{2K}u}H(u+iK') = \sum_{-\infty}^{\infty}(-1)^n q^{n^2} e^{2n\frac{\pi i}{2K}u}.$$

The development of $\Theta(u)$ and further properties are evidently

$$\Theta(u) = 1 - 2q\cos\frac{2\pi u}{2K} + 2q^4\cos\frac{4\pi u}{2K} - 2q^9\cos\frac{6\pi u}{2K} + \cdots, \qquad (24)$$

$$\Theta(u+2K) = \Theta(u), \qquad \Theta(u+2iK') = -q^{-1}e^{-\frac{i\pi}{K}u}\Theta(u). \qquad (25)$$

Finally instead of adding $K + iK'$ to u in $H(u)$, add K in $\Theta(u)$.

$$\Theta_1(u) = 1 + 2q\cos\frac{2\pi u}{2K} + 2q^4\cos\frac{4\pi u}{2K} + 2q^9\cos\frac{6\pi u}{2K} + \cdots, \qquad (26)$$

$$\Theta_1(u+2K) = \Theta_1(u), \qquad \Theta_1(u+2iK') = +q^{-1}e^{-\frac{i\pi}{K}u}\Theta_1(u). \qquad (27)$$

For a tabulation of properties of the four functions see Ex. 1 below.

176. As $H(u)$ vanishes for $u = 0$ and is reproduced except for a finite multiplier when $2mK + 2niK'$ is added to u, the table

$$\begin{aligned} H(u) &= 0 \quad \text{for} \quad u = 2mK + 2niK', \\ H_1(u) &= 0 \quad \text{for} \quad u = (2m+1)K + 2niK', \\ \Theta(u) &= 0 \quad \text{for} \quad u = 2mK + (2n+1)iK', \\ \Theta_1(u) &= 0 \quad \text{for} \quad u = (2m+1)K + (2n+1)iK', \end{aligned}$$

contains the known vanishing points of the four functions. Now it is possible to form infinite products which vanish for these values. From such products it may be seen that the functions have no other vanishing points. Moreover the products themselves are useful.

It will be most convenient to use the function $\Theta_1(u)$. Now

$$e^{\frac{i\pi}{K}(2mK+K+2niK'+iK')} = -q^{(2n+1)}, \qquad -\infty < n < \infty.$$

Hence $$e^{\frac{i\pi}{K}u} + q^{-(2n+1)} \quad \text{and} \quad e^{-\frac{i\pi}{K}u} + q^{-(2n+1)}, \qquad n \geqq 0,$$

are two expressions of which the second vanishes for all the roots of $\Theta_1(u)$ for which $n \geqq 0$, and the first for all roots with $n < 0$. Hence

$$\Pi = C\prod_{1}^{\infty}\left(1 + q^{2n+1}e^{\frac{i\pi}{K}u}\right)\left(1 + q^{2n+1}e^{-\frac{i\pi u}{K}}\right)$$

is an infinite product which vanishes for all the roots of $\Theta_1(u)$. The product is readily seen to converge absolutely and uniformly. In particular it does not diverge to 0 and consequently has no other roots than those of $\Theta_1(u)$ above given. It remains to show that the product is identical with $\Theta_1(u)$ with a proper determination of C.

Let $\Theta_1(u)$ be written in exponential form as follows, with $z = e^{\frac{i\pi}{K}u}$:

$$\phi(z) = \Theta_1(u) = 1 + q\left(z + \frac{1}{z}\right) + q^4\left(z^2 + \frac{1}{z^2}\right) + \cdots + q^{n^2}\left(z^n + \frac{1}{z^n}\right) + \cdots,$$

$$\psi(z) = C^{-1}\prod(u) = (1 + qz)(1 + q^3z)(1 + q^5z)\cdots(1 + q^{2n-1}z)\cdots$$
$$\times\left(1 + \frac{q}{z}\right)\left(1 + \frac{q^3}{z}\right)\left(1 + \frac{q^5}{z}\right)\cdots\left(1 + \frac{q^{2n-1}}{z}\right)\cdots.$$

A direct substitution will show that $\phi(q^2z) = q^{-1}z^{-1}\phi(z)$ and $\psi(q^2z) = q^{-1}z^{-1}\psi(z)$. In fact this substitution is equivalent to replacing u by $u + 2iK'$ in Θ_1. Next consider the first $2n$ terms of $\psi(z)$ written above, and let this finite product be $\psi_n(z)$. Then by substitution

$$(q^{2n} + qz)\psi_n(q^2z) = (1 + q^{2n+1}z)\psi_n(z).$$

Now $\psi_n(z)$ is reciprocal in z in such a way that, if multiplied out,

$$\psi_n(z) = a_0 + a_1\left(z + \frac{1}{z}\right) + a_2\left(z^2 + \frac{1}{z^2}\right) + \cdots + a_n\left(z^n + \frac{1}{z^n}\right), \qquad a_n = q^{n^2}.$$

Then $$(q^{2n} + qz)\sum_0^n a_i(q^{2i}z^i + q^{-2i}z^{-i}) = (1 + q^{2n+1}z)\sum_0^n a_i(z^i + z^{-i}),$$

and the expansion and equation of coefficients of z^i gives the relation

$$a_i = a_{i-1}\frac{q^{2i-1}(1 - q^{2n-2i+2})}{1 - q^{2n+2i}} \quad\text{or}\quad a_i = a_0\frac{q^{i^2}\prod_{k=1}^{i}(1 - q^{2n-2k+2})}{\prod_{k=0}^{i-1}(1 - q^{2n+2k+2})}.$$

From $a_n = q^{n^2}$, $$a_0 = \frac{\prod_{k=0}^{n-1}(1 - q^{2n+2k+2})}{\prod_{k=1}^{n}(1 - q^{2k})}, \qquad a_i = \frac{q^{i^2}\prod_{k=1}^{n-i}(1 - q^{2n+2i+2k})}{\prod_{k=1}^{n-i}(1 - q^{2k})}.$$

Now if n be allowed to become infinite, each coefficient a_i approaches the limit

$$\lim a_i = \frac{q^{i^2}}{C}, \qquad C = \prod_1^\infty(1 - q^{2n}) = (1 - q^2)(1 - q^4)(1 - q^6)\cdots.$$

Hence $$\Theta_1(u) = \prod_1^\infty(1 - q^{2n})\cdot\prod_1^\infty\left(1 + q^{2n-1}e^{\frac{i\pi}{K}u}\right)\left(1 + q^{2n-1}e^{\frac{-i\pi}{K}u}\right),$$

provided the limit of $\psi_n(z)$ may be found by taking the series of the limits of the terms. The justification of this process would be similar to that of § 171.

The products for Θ, H_1, H may be obtained from that for Θ_1 by subtracting K, iK', $K + iK'$ from u and making the needful slight alterations to conform with the definitions. The products may be converted into trigonometric form by multiplying. Then

$$H(u) = C\,2\,q^{\frac{1}{4}}\sin\frac{\pi u}{2K}\prod_1^\infty\left(1 - 2q^{2n}\cos\frac{2\pi u}{2K} + q^{4n}\right). \tag{28}$$

$$H_1(u) = C\, 2\, q^{\frac{1}{4}} \cos \frac{\pi u}{2\,K} \prod_1^{\infty} \left(1 + 2\, q^{2\,n} \cos \frac{2\,\pi u}{2\,K} + q^{4\,n}\right), \tag{29}$$

$$\Theta(u) = C \prod_0^{\infty} \left(1 - 2\, q^{2\,n+1} \cos \frac{2\,\pi u}{2\,K} + q^{4\,n+2}\right), \tag{30}$$

$$\Theta_1(u) = C \prod_0^{\infty} \left(1 + 2\, q^{2\,n+1} \cos \frac{2\,\pi u}{2\,K} + q^{4\,n+2}\right), \tag{31}$$

$$C = \prod_1^{\infty} (1 - q^{2\,n}) = (1 - q^2)\,(1 - q^4)\,(1 - q^6) \cdots, \tag{32}$$

$$H_1(0) = C\, 2\, q^{\frac{1}{4}} \prod_1^{\infty} (1 + q^{2\,n})^2, \qquad \Theta\,(0) = C \prod_0^{\infty} (1 - q^{2\,n+1})^2,$$

$$H'(0) = C\, 2\, q^{\frac{1}{4}} \frac{\pi}{2\,K} \prod_1^{\infty} (1 - q^{2\,n})^2, \qquad \Theta_1(0) = C \prod_0^{\infty} (1 + q^{2\,n+1})^2.$$

The value of $H'(0)$ is found by dividing $H(u)$ by u and letting $u \doteq 0$. Then

$$H'(0) = \frac{\pi}{2\,K} H_1(0)\,\Theta(0)\,\Theta_1(0) \tag{33}$$

follows by direct substitution and cancellation or combination.

177. Other functions may be built from the theta functions. Let

$$\sqrt{k} = \frac{H(K)}{\Theta(K)} = \frac{H_1(0)}{\Theta_1(0)}, \qquad \sqrt{k'} = \frac{\Theta(0)}{\Theta_1(0)}, \qquad \sqrt{\frac{k'}{k}} = \frac{\Theta(0)}{H_1(0)}, \tag{34}$$

$$\operatorname{sn} u = \frac{1}{\sqrt{k}} \frac{H(u)}{\Theta(u)}, \qquad \operatorname{cn} u = \sqrt{\frac{k'}{k}} \frac{H_1(u)}{\Theta(u)}, \qquad \operatorname{dn} u = \sqrt{k'} \frac{\Theta_1(u)}{\Theta(u)}. \tag{35}$$

The functions $\operatorname{sn} u$, $\operatorname{cn} u$, $\operatorname{dn} u$ are called elliptic functions* of u. As H is the only odd theta function, $\operatorname{sn} u$ is odd but $\operatorname{cn} u$ and $\operatorname{dn} u$ are even. *All three functions have two actual periods* in the same sense that $\sin x$ and $\cos x$ have the period $2\,\pi$. Thus $\operatorname{dn} u$ has the periods $2\,K$ and $4\,iK'$ by (25), (27); and $\operatorname{sn} u$ has the periods $4\,K$ and $2\,iK'$ by (25), (21). That $\operatorname{cn} u$ has $4\,K$ and $2\,K + 2\,iK'$ as periods is also easily verified. The values of u which make the functions vanish are known; they are those which make the numerators vanish. In like manner the values of u for which the three functions become infinite are the known roots of $\Theta(u)$.

If q is known, the values of $\sqrt{k}$ and $\sqrt{k'}$ may be found from their definitions. Conversely the expression for $\sqrt{k'}$,

$$\sqrt{k'} = \frac{\Theta(0)}{\Theta_1(0)} = \frac{1 - 2\,q + 2\,q^4 - 2\,q^9 + \cdots}{1 + 2\,q + 2\,q^4 + 2\,q^9 + \cdots}, \tag{36}$$

* The study of the elliptic functions is continued in Chapter XIX.

is readily solved for q by reversion. If powers of q higher than the first are neglected, the approximate value of q is found by solution, as

$$\frac{1}{2}\frac{1-\sqrt{k'}}{1+\sqrt{k'}} = \frac{q+q^9+\cdots}{1-2q^4+\cdots} = q - 2q^5 + 5q^9 + \cdots.$$

Hence $$q = \frac{1}{2}\frac{1-\sqrt{k'}}{1+\sqrt{k'}} + \frac{2}{2^5}\left(\frac{1-\sqrt{k'}}{1+\sqrt{k'}}\right)^5 + \frac{15}{2^9}\left(\frac{1-\sqrt{k'}}{1+\sqrt{k'}}\right)^9 + \cdots \qquad (37)$$

is the series for q. For values of k' near 1 this series converges with great rapidity; in fact if $k'^2 \geqq \frac{1}{2}$, $k' > 0.7$, $\sqrt{k'} > 0.82$, the second term of the expansion amounts to less than $1/10^6$ and may be disregarded in work involving four or five figures. The first two terms here given are sufficient for eleven figures.

Let ϑ denote any one of the four theta series H, H_1, Θ, Θ_1. Then

$$\vartheta^2(u) = \phi(z) = \sum_{-\infty}^{\infty} b_n z^n, \qquad z = e^{\frac{i\pi}{K}u} \qquad (38)$$

may be taken as the form of development of ϑ^2; this is merely the Fourier series for a function with period $2K$. But all the theta functions take the same multiplier, except for sign, when $2iK'$ is added to u; hence the squares of the functions take the same multiplier, and in particular $\phi(q^2z) = q^{-2}z^{-2}\phi(z)$. Apply this relation.

$$\sum b_n q^{2n} z^n = q^{-2}z^{-2}\sum b_n z^n, \qquad b_n q^{2n-2} = b_{n-2}.$$

It then is seen that a recurrent relation between the coefficients is found which will determine all the even coefficients in terms of b_0 and all the odd in terms of b_1. Hence

$$\vartheta^2(u) = b_0\Phi(z) + b_1\Psi(z), \qquad b_0,\ b_1, \text{ constants}, \qquad (38')$$

is the expansion of any ϑ^2 or of any function which may be developed as (38) and satisfies $\phi(q^2z) = q^{-2}z^{-2}\phi(z)$. Moreover Φ and Ψ are identical for all such functions, and the only difference is in the values of the constants b_0, b_1.

As any three theta functions satisfy (38′) with different values of the constants, the functions Φ and Ψ may be eliminated and

$$\alpha\vartheta_1^2(u) + \beta\vartheta_2^2(u) + \gamma\vartheta_3^2(u) = 0,$$

where α, β, γ are constants. In words, the squares of any three theta functions satisfy a linear homogeneous equation with constant coefficients. The constants may be determined by assigning particular values to the argument u. For example, take H, H_1, Θ. Then*

* For brevity the parenthesis about the arguments of a function will frequently be omitted.

$$\alpha H^2(u) + \beta H_1^2(u) = \gamma\Theta^2(u), \qquad \beta H_1^2 0 = \gamma\Theta^2 0, \qquad \alpha H^2 K = \gamma\Theta^2 K,$$

$$\frac{\Theta^2 K}{H^2 K}\frac{H^2(u)}{\Theta^2(u)} + \frac{\Theta^2 0}{H_1^2 0}\frac{H_1^2(u)}{\Theta^2(u)} = 1, \text{ or } \operatorname{sn}^2 u + \operatorname{cn}^2 u = 1. \tag{39}$$

By treating H, Θ_1, Θ in a similar manner may be proved

$$k^2 \operatorname{sn}^2 u + \operatorname{dn}^2 u = 1 \quad \text{and} \quad k^2 + k'^2 = 1. \tag{40}$$

The function $\vartheta(u)\vartheta(u-a)$, where a is a constant, satisfies the relation $\phi(q^2 z) = q^{-2}z^{-2}C\phi(z)$ if $\log C = i\pi a/K$. Reasoning like that used for treating ϑ^2 then shows that between any three such expressions there is a linear relation. Hence

$$\alpha H(u)H(u-a) + \beta H_1(u)H_1(u-a) = \gamma\Theta(u)\Theta(u-a),$$

$$u = 0, \qquad \beta H_1(0)H_1(a) = \gamma\Theta(0)\Theta(a),$$

$$u = K, \qquad \alpha H_1(0)H_1(a) = \gamma\Theta_1(0)\Theta_1(a),$$

$$\frac{\Theta 0\Theta_1 0\Theta_1 a}{H_1^2 0\Theta a}\frac{H(u)H(u-a)}{\Theta(u)\Theta(u-a)} + \frac{\Theta^2 0}{H_1^2 0}\frac{H_1(u)H_1(u-a)}{\Theta(u)\Theta(u-a)} = \frac{\Theta 0}{H_1 0}\frac{H_1 a}{\Theta a},$$

or
$$\operatorname{dn} a \operatorname{sn} u \operatorname{sn}(u-a) + \operatorname{cn} u \operatorname{cn}(u-a) = \operatorname{cn} a. \tag{41}$$

In this relation replace a by $-v$. Then there results

$$\operatorname{cn} u \operatorname{cn}(u+v) + \operatorname{sn} u \operatorname{dn} v \operatorname{sn}(u+v) = \operatorname{cn} v,$$

or
$$\operatorname{cn} v \operatorname{cn}(u+v) + \operatorname{sn} v \operatorname{dn} u \operatorname{sn}(u+v) = \operatorname{cn} u,$$

and
$$\operatorname{sn}(u+v) = \frac{\operatorname{cn}^2 u - \operatorname{cn}^2 v = \operatorname{sn}^2 v - \operatorname{sn}^2 u}{\operatorname{sn} v \operatorname{cn} u \operatorname{dn} u - \operatorname{sn} u \operatorname{cn} v \operatorname{dn} v}, \tag{42}$$

by symmetry and by solution. The fraction may be reduced by multiplying numerator and denominator by the denominator with the middle sign changed, and by noting that

$$\operatorname{sn}^2 v \operatorname{cn}^2 u \operatorname{dn}^2 u - \operatorname{sn}^2 u \operatorname{cn}^2 v \operatorname{dn}^2 v = (\operatorname{sn}^2 v - \operatorname{sn}^2 u)(1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v).$$

Then
$$\operatorname{sn}(u+v) = \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v + \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}, \tag{43}$$

and
$$\operatorname{sn}(u-v) = \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v - \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}.$$

and
$$\operatorname{sn}(u+v) - \operatorname{sn}(u-v) = \frac{2 \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}. \tag{44}$$

The last result may be used to differentiate $\operatorname{sn} u$. For

$$\frac{\operatorname{sn}(u+\Delta u) - \operatorname{sn} u}{\Delta u} = \frac{\operatorname{sn}\frac{1}{2}\Delta u}{\frac{1}{2}\Delta u}\frac{\operatorname{cn}(u+\frac{1}{2}\Delta u)\operatorname{dn}(u+\frac{1}{2}\Delta u)}{1 - k^2\operatorname{sn}^2\frac{1}{2}\Delta u \operatorname{sn}^2(u+\frac{1}{2}\Delta u)}.$$

$$\frac{d}{du}\operatorname{sn} u = g \operatorname{cn} u \operatorname{dn} u, \qquad g = \lim_{u \doteq 0}\frac{\operatorname{sn} u}{u}. \tag{45}$$

Here g is called the *multiplier*. By definition of $\operatorname{sn} u$ and by (33)

$$g = \frac{\Theta_1(0)}{H_1(0)}\frac{H'(0)}{\Theta(0)} = \frac{\pi}{2K}\Theta_1^2(0). \tag{45'}$$

The periods $2K$, $2iK'$ have been independent up to this point. It will, however, be a convenience to have $g=1$ and thus simplify the formula for differentiating $\operatorname{sn} u$. Hence let

$$g = 1, \qquad \sqrt{\frac{2K}{\pi}} = \Theta_1(0) = 1 + 2q + 2q^4 + \cdots. \tag{46}$$

Now of the five quantities K, K', k, k', q only one is independent. If q is known, then k' and K may be computed by (36), (46); k is determined by $k^2 + k'^2 = 1$, and K' by $\pi K'/K = -\log q$ of (19). If, on the other hand, k' is given, q may be computed by (37) and then the other quantities may be determined as before.

EXERCISES

1. With the notations $\lambda = q^{-\frac{1}{4}}e^{-\frac{i\pi}{2K}u}$, $\mu = q^{-1}e^{-\frac{i\pi}{K}u}$ establish:

$$\begin{array}{lll}
H(-u) = -H(u), & H(u+2K) = -H(u), & H(u+2iK') = -\mu H(u),\\
H_1(-u) = +H_1(u), & H_1(u+2K) = -H_1(u), & H_1(u+2iK') = +\mu H_1(u),\\
\Theta(-u) = +\Theta(u), & \Theta(u+2K) = +\Theta(u), & \Theta(u+2iK') = -\mu\Theta(u),\\
\Theta_1(-u) = +\Theta_1(u), & \Theta_1(u+2K) = +\Theta_1(u), & \Theta_1(u+2iK') = +\mu\Theta_1(u),\\
H(u+K) = +H_1(u), & H(u+iK') = i\lambda\Theta(u), & H(u+K+iK') = +\lambda\Theta_1(u),\\
H_1(u+K) = -H(u), & H_1(u+iK') = +\lambda\Theta_1(u), & H_1(u+K+iK') = -i\lambda\Theta(u),\\
\Theta(u+K) = +\Theta_1(u), & \Theta(u+iK') = i\lambda H(u), & \Theta(u+K+iK') = +\lambda H_1(u),\\
\Theta_1(u+K) = +\Theta(u), & \Theta_1(u+iK') = +\lambda H_1(u), & \Theta_1(u+K+iK') = +i\lambda H(u).
\end{array}$$

2. Show that if u is real and $q \leqq \frac{1}{6}$, the first two trigonometric terms in the series for H, H_1, Θ, Θ_1, give four-place accuracy. Show that with $q \leqq 0.1$ these terms give about six-place accuracy.

3. Use $\dfrac{q\sin\alpha}{1-2q\cos\alpha+q^2} = q\sin\alpha + q^2\sin 2\alpha + q^3\sin 3\alpha + \cdots$ to prove

$$\frac{d}{du}\log\Theta(u) = \frac{\Theta'(u)}{\Theta(u)} = \frac{2\pi}{K}\left(\frac{q\sin\frac{\pi u}{K}}{1-q^2} + \frac{q^2\sin\frac{2\pi u}{K}}{1-q^4} + \frac{q^3\sin\frac{3\pi u}{K}}{1-q^6} + \cdots\right).$$

4. Prove the double periodicity of $\operatorname{cn} u$ and show that:

$$\begin{array}{lll}
\operatorname{sn}(u+K) = \dfrac{\operatorname{cn} u}{\operatorname{dn} u}, & \operatorname{sn}(u+iK') = \dfrac{1}{k\operatorname{sn} u}, & \operatorname{sn}(u+K+iK') = \dfrac{\operatorname{dn} u}{k\operatorname{cn} u},\\[2ex]
\operatorname{cn}(u+K) = \dfrac{-k'\operatorname{sn} u}{\operatorname{dn} u}, & \operatorname{cn}(u+iK') = \dfrac{-i\operatorname{dn} u}{k\operatorname{sn} u}, & \operatorname{cn}(u+K+iK') = \dfrac{-ik'}{k\operatorname{cn} u},\\[2ex]
\operatorname{dn}(u+K) = \dfrac{k'}{\operatorname{dn} u}, & \operatorname{dn}(u+iK') = -i\dfrac{\operatorname{cn} u}{\operatorname{sn} u}, & \operatorname{dn}(u+K+iK') = ik'\dfrac{\operatorname{sn} u}{\operatorname{cn} u}.
\end{array}$$

5. Tabulate the values of $\operatorname{sn} u$, $\operatorname{cn} u$, $\operatorname{dn} u$ at 0, K, iK', $K + iK'$.

6. Compute k' and k^2 for $q = \frac{1}{6}$ and $q = 0.1$. Hence show that two trigonometric terms in the theta series give four-place accuracy if $k' \geqq \frac{1}{4}$.

7. Prove $\operatorname{cn}(u+v) = \dfrac{\operatorname{cn} u \operatorname{cn} v - \operatorname{sn} u \operatorname{sn} v \operatorname{dn} u \operatorname{dn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}$,

and
$$\operatorname{dn}(u+v) = \frac{\operatorname{dn} u \operatorname{dn} v - k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn} u \operatorname{cn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}.$$

8. Prove $\dfrac{d}{du} \operatorname{cn} u = - \operatorname{sn} u \operatorname{dn} u$, $\quad \dfrac{d}{du} \operatorname{dn} u = - k^2 \operatorname{sn} u \operatorname{cn} u$, $\quad g = 1$.

9. Prove $\operatorname{sn}^{-1} u = \displaystyle\int_0^u \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}}$ from (45) with $g = 1$.

10. If $g = 1$, compute k, k', K, K', for $q = 0.1$ and $q = 0.01$.

11. If $g = 1$, compute k', q, K, K', for $k^2 = \frac{1}{2}, \frac{3}{4}, \frac{1}{4}$.

12. In Exs. 10, 11 write the trigonometric expressions which give $\operatorname{sn} u$, $\operatorname{cn} u$, $\operatorname{dn} u$ with four-place accuracy.

13. Find $\operatorname{sn} 2u$, $\operatorname{cn} 2u$, $\operatorname{dn} 2u$, and hence $\operatorname{sn} \frac{1}{2} u$, $\operatorname{cn} \frac{1}{2} u$, $\operatorname{dn} \frac{1}{2} u$, and show

$$\operatorname{sn} \tfrac{1}{2} K = (1 + k')^{-\frac{1}{2}}, \qquad \operatorname{cn} \tfrac{1}{2} K = \sqrt{k'}(1+k')^{-\frac{1}{2}}, \qquad \operatorname{dn} \tfrac{1}{2} K = \sqrt{k'}.$$

14. Prove $-k \displaystyle\int \operatorname{sn} u \, du = \log(\operatorname{dn} u + k \operatorname{cn} u)$; also

$$\Theta^2(0) H(u+a) H(u-a) = \Theta^2(a) H^2(u) - H^2(a) \Theta^2(u),$$
$$\Theta^2(0) \Theta(u+a) \Theta(u-a) = \Theta^2(u) \Theta^2(a) - H^2(u) H^2(a).$$

CHAPTER XVIII

FUNCTIONS OF A COMPLEX VARIABLE

178. General theorems. The complex function $u(x, y) + iv(x, y)$, where $u(x, y)$ and $v(x, y)$ are single valued real functions continuous and differentiable partially with respect to x and y, has been defined as a function of the complex variable $z = x + iy$ when and only when the relations $u'_x = v'_y$ and $u'_y = -v'_x$ are satisfied (§ 73). In this case the function has a derivative with respect to z which is independent of the way in which Δz approaches the limit zero. Let $w = f(z)$ be a function of a complex variable. Owing to the existence of the derivative the function is necessarily continuous, that is, if ϵ is an arbitrarily small positive number, a number δ may be found so small that

$$|f(z) - f(z_0)| < \epsilon \quad \text{when} \quad |z - z_0| < \delta, \tag{1}$$

and moreover this relation holds uniformly for all points z_0 of the region over which the function is defined, provided the region includes its bounding curve (see Ex. 3, p. 92).

It is further assumed that the derivatives u'_x, u'_y, v'_x, v'_y are continuous and that therefore the derivative $f'(z)$ is continuous.* The function is then said to be an *analytic function* (§ 126). All the functions of a complex variable here to be dealt with are analytic in general, although they may be allowed to fail of being analytic at certain specified points called *singular points*. The adjective "analytic" may therefore usually be omitted. The equations

$$w = f(z) \quad \text{or} \quad u = u(x, y), \qquad v = v(x, y)$$

define a transformation of the xy-plane into the uv-plane, or, briefer, of the z-plane into the w-plane; to each point of the former corresponds one and only one point of the latter (§ 63). If the Jacobian

$$\begin{vmatrix} u'_x & u'_y \\ v'_x & v'_y \end{vmatrix} = (u'_x)^2 + (v'_x)^2 = |f'(z)|^2 \tag{2}$$

* It may be proved that, in the case of functions of a complex variable, the continuity of the derivative follows from its existence, but the proof will not be given here.

of the transformation does not vanish at a point z_0, the equations may be solved in the neighborhood of that point, and hence to each point of the second plane corresponds only one of the first:

$$x = x(u, v), \qquad y = y(u, v) \quad \text{or} \quad z = \phi(w).$$

Therefore it is seen that *if* $w = f(z)$ *is analytic in the neighborhood of* $z = z_0$, *and if the derivative* $f'(z_0)$ *does not vanish, the function may be solved as* $z = \phi(w)$, where ϕ is the inverse function of f, and is likewise analytic in the neighborhood of the point $w = w_0$. It may readily be shown that, as in the case of real functions, the derivatives $f'(z)$ and $\phi'(w)$ are reciprocals. Moreover, it may be seen that *the transformation is conformal*, that is, that the angle between any two curves is unchanged by the transformation (§ 63). For consider the increments

$$\Delta w = [f'(z_0) + \zeta]\,\Delta z = f'(z_0)\,[1 + \zeta/f'(z_0)]\,\Delta z. \qquad f'(z_0) \neq 0.$$

As Δz and Δw are the chords of the curves before and after transformation, the geometrical interpretation of the equation, apart from the infinitesimal ζ, is that the chords Δz are magnified in the ratio $|f'(z_0)|$ to 1 and turned through the angle of $f'(z_0)$ to obtain the chords Δw (§ 72). In the limit it follows that the tangents to the w-curves are inclined at an angle equal to the angle of the corresponding z-curves plus the angle of $f'(z_0)$. The angle between two curves is therefore unchanged.

The existence of an inverse function and of the geometric interpretation of the transformation as conformal both become illusory at points for which the derivative $f'(z)$ vanishes. Points where $f'(z) = 0$ are called *critical points* of the function (§ 183).

It has further been seen that the integral of a function which is analytic over any simply connected region is independent of the path and is zero around any closed path (§ 124); if the region be not simply connected but the function is analytic, the integral about any closed path which may be shrunk to nothing is zero and the integrals about any two closed paths which may be shrunk into each other are equal (§ 125). Furthermore Cauchy's result that the value

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(t)}{t - z}\,dt \tag{3}$$

of a function, which is analytic upon and within a closed path, may be found by integration around the path has been derived (§ 126). By a transformation the Taylor development of the function has been found whether in the finite form with a remainder (§ 126) or as an infinite series (§ 167). It has also been seen that any infinite power series

which converges is differentiable and hence defines an analytic function within its circle of convergence (§ 166).

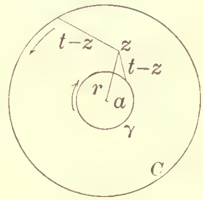
It has also been shown that the sum, difference, product, and quotient of any two functions will be analytic for all points at which both functions are analytic, except at the points at which the denominator, in the case of a quotient, may vanish (Ex. 9, p. 163). The result is evidently extensible to the case of any rational function of any number of analytic functions.

From the possibility of development in series follows that *if two functions are analytic in the neighborhood of a point and have identical values upon any curve drawn through that point*, or even upon any set of points which approach that point as a limit, *then the functions are identically equal within their common circle of convergence and over all regions which can be reached by* (§ 169) *continuing the functions analytically.* The reason is that a set of points converging to a limiting point is all that is needed to prove that two power series are identical provided they have identical values over the set of points (Ex. 9, p. 439). This theorem is of great importance because it shows that if a function is defined for a dense set of real values, any one extension of the definition, which yields a function that is analytic for those values and for complex values in their vicinity, must be equivalent to any other such extension. It is also useful in discussing *the principle of permanence of form;* for if the two sides of an equation are identical for a set of values which possess a point of condensation, say, for all real rational values in a given interval, and if each side is an analytic function, then the equation must be true for all values which may be reached by analytic continuation.

For example, the equation $\sin x = \cos(\frac{1}{2}\pi - x)$ is known to hold for the values $0 \leqq x \leqq \frac{1}{2}\pi$. Moreover the functions $\sin z$ and $\cos z$ are analytic for all values of z whether the definition be given as in § 74 or whether the functions be considered as defined by their power series. Hence the equation must hold for all real or complex values of x. In like manner from the equation $e^xe^y = e^{x+y}$ which holds for real rational exponents, the equation $e^ze^w = e^{z+w}$ holding for all real and imaginary exponents may be deduced. For if y be given any rational value, the functions of x on each side of the sign are analytic for all values of x real or complex, as may be seen most easily by considering the exponential as defined by its power series. Hence the equation holds when x has any complex value. Next consider x as fixed at any desired complex value and let the two sides be considered as functions of y regarded as complex. It follows that the equation must hold for any value of y. The equation is therefore true for any value of z and w.

179. Suppose that a function is analytic in all points of a region except at some one point within the region, and let it be assumed that

the function ceases to be analytic at that point because it ceases to be continuous. The discontinuity may be either finite or infinite. In case the discontinuity is finite let $|f(z)| < G$ in the neighborhood of the point $z = a$ of discontinuity. Cut the point out with a small circle and apply Cauchy's Integral to a ring surrounding the point. The integral is applicable because at all points on and within the ring the function is analytic. If the small circle be replaced by a smaller circle into which it may be shrunk, the value of the integral will not be changed.



$$f(z) = \frac{1}{2\pi i}\left[\int_C \frac{f(t)}{t-z}dt + \int_{\gamma_i} \frac{f(t)}{t-z}dt\right], \qquad i = 1, 2, \cdots.$$

Now the integral about γ_i which is constant can be made as small as desired by taking the circle small enough; for $|f(t)| < G$ and $|t - z| > |a - z| - r_i$, where r_i is the radius of the circle γ_i and hence the integral is less than $2\pi r_i G/[|z - a| - r_i]$. As the integral is constant, it must therefore be 0 and may be omitted. The remaining integral about C, however, defines a function which is analytic at $z = a$. Hence if $f(a)$ be chosen as defined by this integral instead of the original definition, the discontinuity disappears. *Finite discontinuities may therefore be considered as due to bad judgment in defining a function at some point;* and may therefore be disregarded.

In the case of infinite discontinuities, the function may either *become infinite for all methods of approach* to the point of discontinuity, or it may *become infinite for some methods of approach and remain finite for other methods.* In the first case the function is said to have a *pole* at the point $z = a$ of discontinuity; in the second case it is said to have an *essential singularity.* In the case of a pole consider the reciprocal function

$$F(z) = \frac{1}{f(z)}, \qquad z \neq a, \qquad F(a) = 0.$$

The function $F(z)$ is analytic at all points near $z = a$ and remains finite, in fact approaches 0, as z approaches a. As $F(a) = 0$, it is seen that $F(z)$ has no finite discontinuity at $z = a$ and is analytic also at $z = a$. Hence the Taylor expansion

$$F(z) = a_m(z - a)^m + a_{m+1}(z - a)^{m+1} + \cdots$$

is proper. If E denotes a function neither zero nor infinite at $z = a$, the following transformations may be made.

$$F(z) = (z-a)^m E_1(z), \qquad f(z) = (z-a)^{-m} E_2(z),$$

$$f(z) = \frac{C_{-m}}{(z-a)^m} + \frac{C_{-m+1}}{(z-a)^{m-1}} + \cdots + \frac{C_{-1}}{z-a}$$
$$+ C_0 + C_1(z-a) + C_2(z-a)^2 + \cdots.$$

In other words, a function which has a pole at $z = a$ may be written as the product of some power $(z-a)^{-m}$ by an E-function; and as the E-function may be expanded, the function may be expanded into a power series which contains a certain number of negative powers of $(z-a)$. *The order m of the highest negative power* is called *the order of the pole.* Compare Ex. 5, p. 449.

If the function $f(z)$ be integrated around a closed curve lying within the circle of convergence of the series $C_0 + C_1(z-a) + \cdots$, then

$$\int_{\bigcirc} f(z)\,dz = \int \frac{C_{-m}dz}{(z-a)^m} + \cdots + \int_{\bigcirc} \frac{C_{-1}dz}{z-a}$$
$$+ \int_{\bigcirc} [C_0 + C_1(z-a) + \cdots]\,dz = 2\pi i C_{-1},$$

or
$$\int_{\bigcirc} f(z)\,dz = 2\pi i C_{-1}; \tag{4}$$

for the first $m-1$ terms may be integrated and vanish, the term $C_{-1}/(z-a)$ leads to the logarithm $C_{-1}\log(z-a)$ which is multiple valued and takes on the increment $2\pi i C_{-1}$, and the last term vanishes because it is the integral of an analytic function. The total value of the integral of $f(z)$ about a small circuit surrounding a pole is therefore $2\pi i C_{-1}$. The value of the integral about any larger circuit within which the function is analytic except at $z = a$ and which may be shrunk into the small circuit, will also be the same quantity. The coefficient C_{-1} of the term $(z-a)^{-1}$ is called *the residue of the pole;* it cannot vanish if the pole is of the first order, but may if the pole is of higher order.

The discussion of the behavior of a function $f(z)$ when z becomes infinite may be carried on by making a transformation. Let

$$z' = \frac{1}{z}, \qquad z = \frac{1}{z'}, \qquad f(z) = f\left(\frac{1}{z'}\right) = F(z'). \tag{5}$$

To large values of z correspond small values of z'; if $f(z)$ is analytic for all large values of z, then $F(z')$ will be analytic for values of z' near the origin. At $z' = 0$ the function $F(z')$ may not be defined by (5); but if $F(z')$ remains finite for small values of z', a definition may be given so that it is analytic also at $z' = 0$. In this case $F(0)$ is said to be the

value of $f(z)$ when z is infinite and the notation $f(\infty) = F(0)$ may be used. If $F(z')$ does not remain finite but has a pole at $z' = 0$, then $f(z)$ is said to have a pole of the same order at $z = \infty$; and if $F(z')$ has an essential singularity at $z' = 0$, then $f(z)$ is said to have an essential singularity at $z = \infty$. Clearly if $f(z)$ has a pole at $z = \infty$, the value of $f(z)$ must become indefinitely great no matter how z becomes infinite; but if $f(z)$ has an essential singularity at $z = \infty$, there will be some ways in which z may become infinite so that $f(z)$ remains finite, while there are other ways so that $f(z)$ becomes infinite.

Strictly speaking there is no point of the z-plane which corresponds to $z' = 0$. Nevertheless it is convenient to speak as if there were such a point, to call it *the point at infinity*, and to designate it as $z = \infty$. If then $F(z')$ is analytic for $z' = 0$ so that $f(z)$ may be said to be analytic at infinity, the expansions

$$F(z') = C_0 + C_1 z' + C_2 z'^2 + \cdots + C_n z'^n + \cdots =$$

$$f(z) = C_0 + \frac{C_1}{z} + \frac{C_2}{z^2} + \cdots + \frac{C_n}{z^n} + \cdots$$

are valid; the function $f(z)$ has been *expanded about the point at infinity into a descending power series in z*, and the series will converge for all points z *outside* a circle $|z| = R$. For a pole of order m at infinity

$$f(z) = C_{-m} z^m + C_{-m+1} z^{m-1} + \cdots + C_{-1} z + C_0 + \frac{C_1}{z} + \frac{C_2}{z^2} + \cdots.$$

Simply because it is convenient to introduce the concept of the point at infinity for the reason that in many ways the totality of large values for z does not differ from the totality of values in the neighborhood of a finite point, it should not be inferred that the point at infinity has all the properties of finite points.

EXERCISES

1. Discuss $\sin(x + y) = \sin x \cos y + \cos x \sin y$ for permanence of form.

2. If $f(z)$ has an essential singularity at $z = a$, show that $1/f(z)$ has an essential singularity at $z = a$. Hence infer that there is some method of approach to $z = a$ such that $f(z) \doteq 0$.

3. By treating $f(z) - c$ and $[f(z) - c]^{-1}$ show that at an essential singularity a function may be made to approach any assigned value c by a suitable method of approaching the singular point $z = a$.

4. Find the order of the poles of these functions at the origin:

(α) $\cot z$, $\quad$ (β) $\csc^2 z \log(1 - z)$, $\quad$ (γ) $z(\sin z - \tan z)^{-1}$.

5. Show that if $f(z)$ vanishes at $z = a$ once or n times, the quotient $f'(z)/f(z)$ has the residue 1 or n. Show that if $f(z)$ has a pole of the mth order at $z = a$, the quotient has the residue $-m$.

6. From Ex. 5 prove the important theorem that: If $f(z)$ is analytic and does not vanish upon a closed curve and has no singularities other than poles within the curve, then

$$\frac{1}{2\pi i}\int_{\bigcirc}\frac{f'(z)}{f(z)}dz = n_1 + n_2 + \cdots + n_k - m_1 - m_2 - \cdots - m_l = N - M,$$

where N is the total number of roots of $f(z) = 0$ within the curve and M is the sum of the orders of the poles.

7. Apply Ex. 6 to $1/P(z)$ to show that a polynomial $P(z)$ of the nth order has just n roots within a sufficiently large curve.

8. Prove that e^z cannot vanish for any finite value of z.

9. Consider the residue of $zf'(z)/f(z)$ at a pole or vanishing point of $f(z)$. In particular prove that if $f(z)$ is analytic and does not vanish upon a closed curve and has no singularities but poles within the curve, then

$$\frac{1}{2\pi i}\int_{\bigcirc}\frac{zf'(z)}{f(z)}dz = n_1a_1 + n_2a_2 + \cdots + n_ka_k - m_1b_1 - m_2b_2 - \cdots - m_lb_l,$$

where $a_1, a_2, \cdots, a_k$ and $n_1, n_2, \cdots, n_k$ are the positions and orders of the roots, and $b_1, b_2, \cdots, b_l$ and $m_1, m_2, \cdots, m_l$ of the poles of $f(z)$.

10. Prove that $\Theta_1(z)$, p. 469, has only one root within a rectangle $2K$ by $2iK'$.

11. State the behavior (analytic, pole, or essential singularity) at $z = \infty$ for:

$$(\alpha)\ z^2 + 2z, \qquad (\beta)\ e^{z}, \qquad (\gamma)\ z/(1+z), \qquad (\delta)\ z/(z^3+1).$$

12. Show that if $f(z) = (z-a)^kE(z)$ with $-1 < k < 0$, the integral of $f(z)$ about an infinitesimal contour surrounding $z = a$ is infinitesimal. What analogous theorem holds for an infinite contour?

180. Characterization of some functions. The study of the limitations which are put upon a function when certain of its properties are known is important. For example, *a function which is analytic for all values of z including also $z = \infty$ is a constant.* To show this, note that as the function nowhere becomes infinite, $|f(z)| < G$. Consider the difference $f(z_0) - f(0)$ between the value at any point $z = z_0$ and at the origin. Take a circle concentric with $z = 0$ and of radius $R > |z_0|$. Then by Cauchy's Integral

$$f(z_0) - f(0) = \frac{1}{2\pi i}\left[\int\frac{f(t)}{t - z_0}dt - \int\frac{f(t)}{t-0}dt\right] = \frac{z_0}{2\pi i}\int_{C}\frac{f(t)dt}{t(t-z_0)},$$

or
$$|f(z_0) - f(0)| < \frac{|z_0|}{2\pi}\frac{2\pi RG}{R(R - |z_0|)} = \frac{G|z_0|}{R - |z_0|}.$$

By taking R large enough the difference, which is constant, may be made as small as desired and hence must be zero; hence $f(z) = f(0)$.

Any rational function $f(z) = P(z)/Q(z)$, where $P(z)$ and $Q(z)$ are polynomials in z and may be assumed to be devoid of common factors, can have as singularities merely poles. There will be a pole at each point at which the denominator vanishes; and if the degree of the numerator exceeds that of the denominator, there will be a pole at infinity of order equal to the difference of those degrees. Conversely it may be shown that *any function which has no other singularity than a pole of the mth order at infinity must be a polynomial of the mth order:* that *if the only singularities are a finite number of poles, whether at infinity or at other points, the function is a rational function;* and finally that *the knowledge of the zeros and poles with the multiplicity or order of each is sufficient to determine the function except for a constant multiplier.*

For, in the first place, if $f(z)$ is analytic except for a pole of the mth order at infinity, the function may be expanded as

$$f(z) = a_{-m}z^m + \cdots + a_{-1}z + a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots,$$

or $$f(z) - [a_{-m}z^m + \cdots + a_{-1}z] = a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots.$$

The function on the right is analytic at infinity, and so must its equal on the left be. The function on the left is the difference of a function which is analytic for all finite values of z and a polynomial which is also analytic for finite values. Hence the function on the left or its equal on the right is analytic for all values of z including $z = \infty$, and is a constant, namely a_0. Hence

$$f(z) = a_0 + a_{-1}z + \cdots + a_{-m}z^m \quad \text{is a polynomial of order } m.$$

In the second place let $z_1, z_2, \cdots, z_k, \infty$ be poles of $f(z)$ of the respective orders $m_1, m_2, \cdots, m_k, m$. The function

$$\phi(z) = (z - z_1)^{m_1}(z - z_2)^{m_2} \cdots (z - z_k)^{m_k} f(z)$$

will then have no singularity but a pole of order $m_1 + m_2 + \cdots + m_k + m$ at infinity; it will therefore be a polynomial, and $f(z)$ is rational. As the numerator $\phi(z)$ of the fraction cannot vanish at $z_1, z_2, \cdots, z_k$, but must have $m_1 + m_2 + \cdots + m_k + m$ roots, the knowledge of these roots will determine the numerator $\phi(z)$ and hence $f(z)$ except for a constant multiplier. It should be noted that if $f(z)$ has not a pole at infinity but has a zero of order m, the above reasoning holds on changing m to $-m$.

When $f(z)$ has a pole at $z = a$ of the mth order, the expansion of $f(z)$ about the pole contains certain negative powers

$$P(z - a) = \frac{c_{-m}}{(z - a)^m} + \frac{c_{-m+1}}{(z - a)^{m-1}} + \cdots + \frac{c_{-1}}{z - a}$$

and the difference $f(z) - P(z - a)$ is analytic at $z = a$. The terms $P(z - a)$ are called *the principal part of the function $f(z)$ at the pole a.*

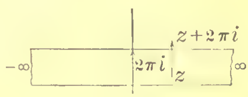
If the function has only a finite number of finite poles and the principal parts corresponding to each pole are known,

$$\phi(z) = f(z) - P_1(z - z_1) - P_2(z - z_2) - \cdots - P_k(z - z_k)$$

is a function which is everywhere analytic for finite values of z and behaves at $z = \infty$ just as $f(z)$ behaves there, since $P_1, P_2, \cdots, P_k$ all vanish at $z = \infty$. If $f(z)$ is analytic at $z = \infty$, then $\phi(z)$ is a constant; if $f(z)$ has a pole at $z = \infty$, then $\phi(z)$ is a polynomial in z and all of the polynomial except the constant term is the principal part of the pole at infinity. Hence *if a function has no singularities except a finite number of poles, and the principal parts at these poles are known, the function is determined except for an additive constant.*

From the above considerations it appears that if a function has no other singularities than a finite number of poles, the function is rational; and that, moreover, the function is determined in factored form, except for a constant multiplier, when the positions and orders of the finite poles and zeros are known; or is determined, except for an additive constant, in a development into partial fractions if the positions and principal parts of the poles are known. All single valued functions other than rational functions must therefore have either an infinite number of poles or some essential singularities.

181. The exponential function $e^z = e^x(\cos y + i \sin y)$ has no finite singularities and its singularity at infinity is necessarily essential. The function is periodic (§ 74) with the period $2\pi i$, and hence will take on all the different values which it can have, if z, instead of being allowed all values, is restricted to have its pure imaginary part y between two limits $y_0 \leqq y < y_0 + 2\pi$; that is, to consider the values of e^z it is merely necessary to consider the values in a strip of the z-plane parallel to the axis of reals and of breadth 2π (but lacking one edge). For convenience the strip may be taken immediately above the axis of reals. The function e^z becomes infinite as z moves out toward the right, and zero as z moves out toward the left in the strip. If $c = a + bi$ is any number other than 0, there is one and only one point in the strip at which $e^z = c$. For



$$e^x = \sqrt{a^2 + b^2} \quad \text{and} \quad \cos y + i \sin y = \frac{a}{\sqrt{a^2 + b^2}} + i \frac{b}{\sqrt{a^2 + b^2}}$$

have only one solution for x and only one for y if y be restricted to an interval 2π. All other points for which $e^z = c$ have the same value for x and some value $y \pm 2n\pi$ for y.

Any rational function of e^z, as

$$R(e^z) = C\,\frac{e^{nz} + a_1 e^{(n-1)z} + \cdots + a_{n-1}e^z + a_n}{e^{mz} + b_1 e^{(m-1)z} + \cdots + b_{m-1}e^z + b_m},$$

will also have the period $2\pi i$. When z moves off to the left in the strip, $R(e^z)$ will approach Ca_n/b_m if $b_m \neq 0$ and will become infinite if $b_m = 0$. When z moves off to the right, $R(e^z)$ must become infinite if $n > m$, approach C if $n = m$, and approach 0 if $n < m$. The denominator may be factored into terms of the form $(e^z - \alpha)^k$, and if the fraction is in its lowest terms each such factor will represent a pole of the kth order in the strip because $e^z - \alpha = 0$ has just one simple root in the strip. Conversely it may be shown that: *Any function $f(z)$ which has the period $2\pi i$, which further has no singularities but a finite number of poles in each strip, and which either becomes infinite or approaches a finite limit as z moves off to the right or to the left, must be $f(z) = R(e^z)$, a rational function of e^z.*

The proof of this theorem requires several steps. Let it first be assumed that $f(z)$ remains finite at the ends of the strip and has no poles. Then $f(z)$ is finite over all values of z, including $z = \infty$, and must be merely constant. Next let $f(z)$ remain finite at the ends of the strip but let it have poles at some points in the strip. It will be shown that a rational function $R(e^z)$ may be constructed such that $f(z) - R(e^z)$ remains finite all over the strip, including the portions at infinity, and that therefore $f(z) = R(e^z) + C$. For let the principal part of $f(z)$ at any pole $z = c$ be

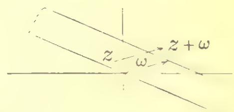
$$P(z-c) = \frac{c_{-k}}{(z-c)^k} + \frac{c_{-k+1}}{(z-c)^{k-1}} + \cdots + \frac{c_{-1}}{z-c}; \quad \text{then} \quad \frac{c_{-k}e^{kc}}{(e^z - e^c)^k} - \frac{c_{-k}}{(z-c)^k} + \cdots$$

is a rational function of e^z which remains finite at both ends of the strip and is such that the difference between it and $P(z-c)$ or $f(z)$ has a pole of not more than the $(k-1)$st order at $z = c$. By subtracting a number of such terms from $f(z)$ the pole at $z = c$ may be eliminated without introducing any new pole. Thus all the poles may be eliminated, and the result is proved.

Next consider the case where $f(z)$ becomes infinite at one or at both ends of the strip. If $f(z)$ happens to approach 0 at one end, consider $f(z) + C$, which cannot approach 0 at either end of the strip. Now if $f(z)$ or $f(z) + C$, as the case may be, had an infinite number of zeros in the strip, these zeros would be confined within finite limits and would have a point of condensation and the function would vanish identically. It must therefore be that the function has only a finite number of zeros; its reciprocal will therefore have only a finite number of poles in the strip and will remain finite at the ends of the strips. Hence the reciprocal and consequently the function itself is a rational function of e^z. The theorem is completely demonstrated.

If the relation $f(z + \omega) = f(z)$ is satisfied by a function, the function is said to have the period ω. The function $f(2\pi iz/\omega)$ will then have the period $2\pi i$. Hence it follows that *if $f(z)$ has the period ω, becomes infinite or remains finite at the ends of a strip of vector breadth*

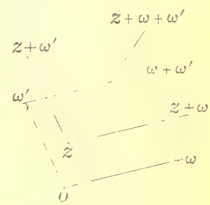
ω, *and has no singularities but a finite number of poles in the strip, the function is a rational function of* $e^{2\pi iz/\omega}$. In particular if the period is 2π, the function is rational in e^{iz}, as is the case with $\sin z$ and $\cos z$; and if the period is π, the function is rational in $e^{iz/2}$, as is $\tan z$. It thus appears that the single valued elementary functions, namely, rational functions, and rational functions of the exponential or trigonometric functions, have simple general properties which are characteristic of these classes of functions.



182. Suppose a function $f(z)$ has two independent periods so that

$$f(z+\omega)=f(z), \qquad f(z+\omega')=f(z).$$

The function then has the same value at z and at any point of the form $z+m\omega+n\omega'$, where m and n are positive or negative integers. The function takes on all the values of which it is capable in a parallelogram constructed on the vectors ω and ω'. Such a function is called *doubly periodic*. As the values of the function are the same on opposite sides of the parallelogram, only two sides and the one included vertex are supposed to belong to the figure. It has been seen that some doubly periodic functions exist (§ 177); but without reference to these special functions many important theorems concerning doubly periodic functions may be proved, subject to a subsequent demonstration that the functions do exist.



If a doubly periodic function has no singularities in the parallelogram, it must be constant; for the function will then have no singularities at all. *If two periodic functions have the same periods and have the same poles and zeros* (each to the same order) *in the parallelogram, the quotient of the functions is a constant: if they have the same poles and the same principal parts at the poles, their difference is a constant.* In these theorems (and all those following) it is assumed that the functions have no essential singularity in the parallelogram. The proof of the theorems is left to the reader. If $f(z)$ is doubly periodic, $f'(z)$ is also doubly periodic. The integral of a doubly periodic function taken around any parallelogram equal and parallel to the parallelogram of periods is zero; for the function repeats itself on opposite sides of the figure while the differential dz changes sign. Hence in particular

$$\int_{\square} f(z)\,dz=0, \qquad \int_{\square}\frac{f'(z)}{f(z)}\,dz=0, \qquad \int_{\square}\frac{f'(z)\,dz}{f(z)-C}=0.$$

The first integral shows that *the sum of the residues of the poles in the parallelogram is zero;* the second, that *the number of zeros is equal to the number of poles* provided multiplicities are taken into account; the third, that *the number of zeros of $f(z) - C$ is the same as the number of zeros or poles of $f(z)$*, because the poles of $f(z)$ and $f(z) - C$ are the same.

The common number m of poles of $f(z)$ or of zeros of $f(z)$ or of roots of $f(z) = C$ in any one parallelogram is called *the order of the doubly periodic function.* As the sum of the residues vanishes, it is impossible that there should be a single pole of the first order in the parallelogram. Hence there can be no functions of the first order and the simplest possible functions would be of the second order with the expansions

$$\frac{1}{(z-a)^2} + c_0 + c_1(z-c) + \cdots \quad \text{or} \quad \frac{1}{z-a_1} + c_0 + \cdots \quad \text{and} \quad \frac{-1}{z-a_2} + c'_0 + \cdots$$

in the neighborhood of a single pole at $z = a$ of the second order or of the two poles of the first order at $z = a_1$ and $z = a_2$. Let it be assumed that when the periods ω, ω' are given, a doubly periodic function $g(z, a)$ with these periods and with a double pole at $z = a$ exists, and similarly that $h(z, a_1, a_2)$ with simple poles at a_1 and a_2 exists.

Any doubly periodic function $f(z)$ with the periods ω, ω' may be expressed as a polynomial in the functions $g(z, a)$ and $h(z, a_1, a_2)$ of the second order. For in the first place if the function $f(z)$ has a pole of even order $2k$ at $z = a$, then $f(z) - C[g(z, a)]^k$, where C is properly chosen, will have a pole of order less than $2k$ at $z = a$ and will have no other poles than $f(z)$. Hence the order of $f(z) - C[g(z, a)]^k$ is less than that of $f(z)$. And if $f(z)$ has a pole of odd order $2k + 1$ at $z = a$, the function $f(z) - C[g(z, a)]^k h(z, a, b)$, with the proper choice of C, will have a pole of order $2k$ or less at $z = a$ and will gain a simple pole at $z = b$. Thus although $f - Cg^k h$ will generally not be of lower order than f, it will have a complex pole of odd order split into a pole of even order and a pole of the first order; the order of the former may be reduced as before and pairs of the latter may be removed. By repeated applications of the process a function may be obtained which has no poles and must be constant. The theorem is therefore proved.

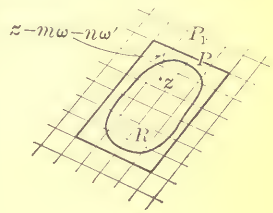
With the aid of series it is possible to write down some doubly periodic functions. In particular consider the series

$$p(z) = \frac{1}{z^2} + \sum{}' \left[\frac{1}{(z - m\omega - n\omega')^2} - \frac{1}{(m\omega + n\omega')^2} \right] \tag{6}$$

and
$$p'(z) = -2 \sum \frac{1}{(z - m\omega - n\omega')^3},$$

where the second Σ denotes summation extended over all values of m, n, whether positive or negative or zero, and Σ' denotes summation extended over all these values except the pair $m = n = 0$. As the summations extend over all possible values for m, n, the series constructed for $z + \omega$ and for $z + \omega'$ must have the same terms as those for z, the only difference being a different arrangement of the terms. If, therefore, the series are absolutely convergent so that the order of the terms is immaterial, the functions must have the periods ω, ω'.

Consider first the convergence of the series $\wp'(z)$. For $z = m\omega + n\omega'$, that is, at the vertices of the net of parallelograms one term of the series becomes infinite and the series cannot converge. But if z be restricted to a finite region R about $z = 0$, there will be only a finite number of terms which can become infinite. Let a parallelogram P large enough to surround the region be drawn, and consider only the vertices which lie outside this parallelogram. For convenience of computation let the points $z = m\omega + n\omega'$ outside P be considered as arranged on successive parallelograms $P_1, P_2, \cdots, P_k, \cdots$. If the number of vertices on P be ν, the number on P_1 is $\nu + 8$ and on P_k is $\nu + 8k$. The shortest vector $z - m\omega - n\omega'$ from z to any vertex of P_1 is longer than a, where a is the least altitude of the parallelogram of periods. The total contribution of P_1 to $\wp'(z)$ is therefore less than $(\nu + 8)a^{-3}$ and the value contributed by all the vertices on successive parallelograms will be less than



$$S = \frac{\nu + 8}{a^3} + \frac{\nu + 8 \cdot 2}{(2a)^3} + \frac{\nu + 8 \cdot 3}{(3a)^3} + \cdots + \frac{\nu + 8 \cdot k}{(ka)^3} + \cdots.$$

This series of positive terms converges. Hence the infinite series for $\wp'(z)$, when the first terms corresponding to the vertices within P_1 are disregarded, converges absolutely and even uniformly so that it represents an analytic function. The whole series for $\wp'(z)$ therefore represents a doubly periodic function of the *third order* analytic everywhere except at the vertices of the parallelograms where it has a pole of the third order. As the part of the series $\wp'(z)$ contributed by vertices outside P is uniformly convergent, it may be integrated from 0 to z to give the corresponding terms in $\wp(z)$ which will also be absolutely convergent because the terms, grouped as for $\wp'(z)$, will be less than the terms of lS where l is the length of the path of integration from 0 to z. The other terms of $\wp'(z)$, thus far disregarded, may be integrated at sight to obtain the corresponding terms of $\wp(z)$. Hence $\wp'(z)$ is really the derivative of $\wp(z)$; and as $\wp(z)$ converges absolutely except for the vertices of the parallelograms, it is clearly doubly periodic of the *second order* with the periods ω, ω', for the same reason that $\wp'(z)$ is periodic.

It has therefore been shown that doubly periodic functions exist, and hence the theorems deduced for such functions are valid. Some further important theorems are indicated among the exercises. They lead to the inference that any doubly periodic function which has the

periods ω, ω' and has no other singularities than poles may be expressed as a rational function of $p(z)$ and $p'(z)$, or as an irrational function of $p(z)$ alone, the only irrationalities being square roots. Thus by employing only the general methods of the theory of functions of a complex variable an entirely new category of functions has been characterized and its essential properties have been proved.

EXERCISES

1. Find the principal parts at $z=0$ for the functions of Ex. 4, p. 481.

2. Prove by Ex. 6, p. 482, that $e^z - e = 0$ has only one root in the strip.

3. How does $e^{(e^z)}$ behave as z becomes infinite in the strip?

4. If the values $R(e^z)$ approaches when z becomes infinite in the strip are called exceptional values, show that $R(e^z)$ takes on every value other than the exceptional values k times in the strip, k being the greater of the two numbers n, m.

5. Show by Ex. 9, p. 482, that in any parallelogram of periods the sum of the positions of the roots less the sum of the positions of the poles of a doubly periodic function is $m\omega + n\omega'$, where m and n are integers.

6. Show that the terms of $p'(z)$ may be associated in such a way as to prove that $p'(-z) = -p'(z)$, and hence infer that the expansions are

$$p'(z) = -2z^{-3} + 2c_1z + 4c_2z^3 + \cdots, \qquad \text{only odd powers,}$$

and

$$p(z) = z^{-2} + c_1z^2 + c_2z^4 + \cdots, \qquad \text{only even powers.}$$

7. Examine the series (6) for $p'(z)$ to show that $p'(\frac{1}{2}\omega) = p'(\frac{1}{2}\omega') = p'(\frac{1}{2}\omega + \frac{1}{2}\omega') = 0$. Why can $p'(z)$ not vanish for any other points in the parallelogram?

8. Let $p(\frac{1}{2}\omega) = e$, $p(\frac{1}{2}\omega') = e'$, $p(\frac{1}{2}\omega + \frac{1}{2}\omega') = e''$. Prove the identity of the doubly periodic functions $[p'(z)]^2$ and $4[p(z) - e][p(z) - e'][p(z) - e'']$.

9. By examining the series defining $p(z)$ show that any two points $z = a$ and $z = a'$ such that $p(a) = p(a')$ are symmetrically situated in the parallelogram with respect to the center $z = \frac{1}{2}(\omega + \omega')$. How could this be inferred from Ex. 5?

10. With the notations $g(z, a)$ and $h(z, a_1, a_2)$ of the text show:

$$(\alpha)\quad \frac{p'(z) + p'(a)}{p(z) - p(a)} = 2h(z, 0, a), \qquad \frac{p'(z) + p'(a)}{p(z) - p(a)} = -2h(z, a, 0),$$

$$(\beta)\quad \frac{p'(z) + p'(a_2)}{p(z) - p(a_2)} - \frac{p'(z) + p'(a_1)}{p(z) - p(a_1)} = 2h(z, a_1, a_2),$$

$$(\gamma)\quad \frac{1}{4}\left[\frac{p'(z) + p'(a)}{p(z) - p(a)}\right]^2 - p(z) = g(z, a) = p(z - a) + \text{const.},$$

$$(\delta)\quad p(z - a) = \frac{1}{4}\left[\frac{p'(z) + p'(a)}{p(z) - p(a)}\right]^2 - p(z) - p(a).$$

11. Demonstrate the final theorem of the text of § 182.

12. By combining the power series for $p(z)$ and $p'(z)$ show

$$[p'(z)]^2 - 4[p(z)]^3 + 20c_1\, p(z) + 28c_2 = Az^2 + \text{higher powers.}$$

Hence infer that the right-hand side must be identically zero.

13. Combine Ex. 12 with Ex. 8 to prove $e + e' + e'' = 0$.

14. With the notations $g_2 = 20c_1$ and $g_3 = 28c_2$ show

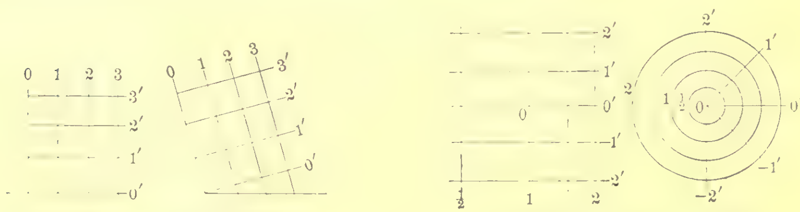
$$p'(z) = \sqrt{4p^3(z) - g_2 p(z) - g_3} \quad \text{or} \quad \frac{dp}{\sqrt{4p^3 - g_2 p - g_3}} = dz.$$

15. If $\zeta(z)$ be defined by $-\dfrac{d}{dz}\zeta(z) = p(z)$ or $\zeta(z) = -\int p(z)dz$, show that $\zeta(z + \omega) - \zeta(z)$ and $\zeta(z + \omega') - \zeta(z)$ must be merely constants η and η'.

183. Conformal representation. The transformation (§ 178)

$$w = f(z) \quad \text{or} \quad u + iv = u(x, y) + iv(x, y)$$

is conformal between the planes of z and w at all points z at which $f'(z) \neq 0$. The correspondence between the planes may be represented by ruling the z-plane and drawing the corresponding rulings in the w-plane. If in particular the rulings in the z-plane be the lines $x =$ const., $y =$ const., parallel to the axes, those in the w-plane must be two sets of curves which are also orthogonal; in like manner if the z-plane be ruled by circles concentric with the origin and rays issuing from the origin, the w-plane must also be ruled orthogonally; for in both cases the angles between curves must be preserved. It is usually most convenient to consider the w-plane as ruled with the lines $u =$ const., $v =$ const., and hence to have a set of rulings $u(x, y) = c_1$, $v(x, y) = c_2$ in the z-plane. The figures represent several different cases arising from the functions:



w–*plane* (1) *z*–*plane* *w*–*plane* (2) *z*–*plane*

(1) $w = az = (a_1 + a_2 i)(x + iy), \qquad u = a_1 x - a_2 y, \qquad v = a_2 x + a_1 y.$

(2) $w = \log z = \log \sqrt{x^2 + y^2} + i \tan^{-1}\dfrac{y}{x}, \quad u = \log \sqrt{x^2 + y^2}, \quad v = \tan^{-1}\dfrac{y}{x}.$

Consider $w = z^2$, and apply polar coördinates so that

$$w = R(\cos\Phi + i\sin\Phi) = r^2(\cos 2\phi + i\sin 2\phi), \qquad R = r^2, \qquad \Phi = 2\phi.$$

To any point (r, ϕ) in the z-plane corresponds $(R = r^2, \Phi = 2\phi)$ in the w-plane; circles about $z = 0$ become circles about $w = 0$ and rays issuing from $z = 0$ become rays issuing from $w = 0$ at twice the angle. (A figure to scale should be supplied by the reader.) The derivative $w' = 2z$ vanishes at $z = 0$ only. The transformation is conformal for all points except $z = 0$. At $z = 0$ it is clear that the angle between two curves in the z-plane is doubled on passing to the corresponding curves in the w-plane; hence at $z = 0$ the transformation is not conformal. Similar results would be obtained from $w = z^m$ except that the angle between rays issuing from $w = 0$ would be m times the angle between the rays at $z = 0$.

A point in the neighborhood of which a function $w = f(z)$ is analytic but has a vanishing derivative $f'(z)$ is called a *critical point* of $f(z)$; if the derivative $f'(z)$ has a root of multiplicity k at any point, that point is called a *critical point of order* k. Let $z = z_0$ be a critical point of order k. Expand $f'(z)$ as

$$f'(z) = a_k(z - z_0)^k + a_{k+1}(z - z_0)^{k+1} + a_{k+2}(z - z_0)^{k+2} + \cdots;$$

then $$f(z) = f(z_0) + \frac{a_k}{k+1}(z - z_0)^{k+1} + \frac{a_{k+1}}{k+2}(z - z_0)^{k+2} + \cdots,$$

or $$w = w_0 + (z - z_0)^{k+1}E(z) \quad \text{or} \quad w - w_0 = (z - z_0)^{k+1}E(z), \qquad (7)$$

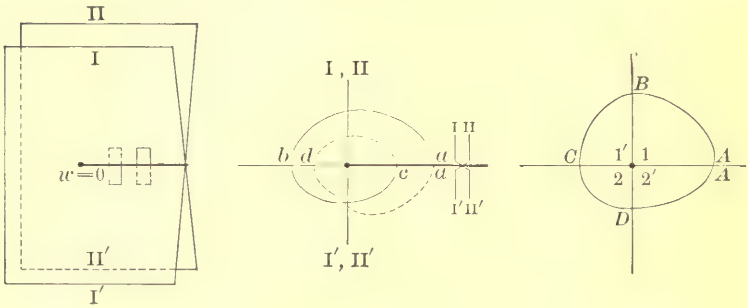
where E is a function that does not vanish at z_0. The point $z = z_0$ goes into $w = w_0$. For a sufficiently small region about z_0 the transformation (7) is sufficiently represented as

$$w - w_0 = C(z - z_0)^{k+1}, \qquad C = E(z_0).$$

On comparison with the case $w = z^m$, it appears that the angle between two curves meeting at z_0 will be multiplied by $k + 1$ on passing to the corresponding curves meeting at w_0. Hence *at a critical point of the kth order the transformation is not conformal but angles are multiplied by $k + 1$ on passing from the z-plane to the w-plane.*

Consider the transformation $w = z^2$ more in detail. To each point z corresponds one and only one point w. To the points z in the first quadrant correspond the points of the first two quadrants in the w-plane, and to the upper half of the z-plane corresponds the whole w-plane. In like manner the lower half of the z-plane will be mapped upon the whole w-plane. Thus in finding the points in the w-plane which correspond to all the points of the z-plane, the w-plane is covered twice. This double counting of the w-plane may be obviated by a simple device. Instead of having one sheet of paper to represent the w-plane,

let two sheets be superposed, and let the points corresponding to the upper half of the z-plane be considered as in the upper sheet, while those corresponding to the lower half are considered as in the lower sheet. Now consider the path traced upon the double w-plane when z traces a path in the z-plane. Every time z crosses from the second to



w—surface *w—surface* *z—plane*

the third quadrant, w passes from the fourth quadrant of the upper sheet into the first of the lower. When z passes from the fourth to the first quadrants, w comes from the fourth quadrant of the lower sheet into the first of the upper.

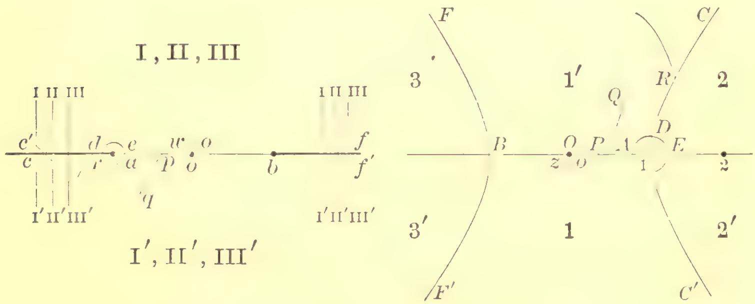
It is convenient to join the two sheets into a single surface so that a continuous path on the z-plane is pictured as a continuous path on the w-surface. This may be done (as indicated at the right of the middle figure) by regarding the lower half of the upper sheet as connected to the upper half of the lower, and the lower half of the lower as connected to the upper half of the upper. The surface therefore cuts through itself along the positive axis of reals, as in the sketch on the left*; the line is called the *junction line* of the surface. The point $w = 0$ which corresponds to the critical point $z = 0$ is called the *branch point* of the surface. Now not only does one point of the z-plane go over into a single point of the w-surface, but to each point of the surface corresponds a single point z; although any two points of the w-surface which are superposed have the same value of w, they correspond to different values of z except in the case of the branch point.

184. The w-surface, which has been obtained as a mere convenience in mapping the z-plane on the w-plane, is of particular value in studying the inverse function $z = \sqrt{w}$. For $\sqrt{w}$ is a multiple valued function and to each value of w correspond two values of z; but if w be

* Practically this may be accomplished for two sheets of paper by pasting gummed strips to the sheets which are to be connected across the cut.

regarded as on the w-surface instead of merely in the w-plane, there is only one value of z corresponding to a point w upon the surface. *Thus the function $\sqrt{w}$ which is double valued over the w-plane becomes single valued over the w-surface.* The w-surface is called the *Riemann surface* of the function $z = \sqrt{w}$. The construction of Riemann surfaces is important in the study of multiple valued functions because the surface keeps the different values apart, so that to each point of the surface corresponds only one value of the function. Consider some surfaces. (The student should make a paper model by following the steps as indicated.)

Let $w = z^3 - 3z$ and plot the w-surface. First solve $f'(z) = 0$ to find the critical points z and substitute to find the branch points w. Now if the branch points be considered as removed from the w-plane, the plane is no longer simply connected. It must be made simply connected by drawing proper lines in the figure. This may be accomplished by drawing a line from each branch point to infinity or by connecting the successive branch points to each other and connecting the last one to the point at infinity. These lines are the junction lines. In this particular case the critical points are $z = +1, -1$ and the branch points are $w = -2, +2$, and the junction lines may be taken as the straight lines joining $w = -2$ and $w = +2$ to



w-surface *z-plane*

infinity and lying along the axis of reals as in the figure. Next spread the requisite number of sheets over the w-plane and cut them along the junction lines. As $w = z^3 - 3z$ is a cubic in z, and to each value of w, except the branch values, there correspond three values of z, three sheets are needed. Now find in the z-plane the *image* of the junction lines. The junction lines are represented by $v = 0$; but $v = 3x^2y - y^3 - 3y$, and hence the line $y = 0$ and the hyperbola $3x^2 - y^2 = 3$ will be the images desired. The z-plane is divided into six pieces which will be seen to correspond to the six half sheets over the w-plane.

Next z will be made to trace out the images of the junction lines and to turn about the critical points so that w will trace out the junction lines and turn about the branch points in such a manner that the connections between the different sheets may be made. It will be convenient to regard z and w as persons walking along their respective paths so that the terms "right" and "left" have a meaning.

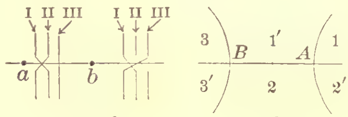
Let z start at $z = 0$ and move forward to $z = 1$; then, as $f'(z)$ is negative, w starts at $w = 0$ and moves back to $w = -2$. Moreover if z turns to the right as at P, so must w turn to the right through the same angle, owing to the conformal property. Thus it appears that not only is OA mapped on oa, but the region 1′ just above OA is mapped on the region I′ just below oa; in like manner OB is mapped on ob. As ab is not a junction line and the sheets have not been cut through along it, the regions 1, 1′ should be assumed to be mapped on the same sheet, say, the uppermost, I, I′. As any point Q in the whole infinite region 1′ may be reached from 0 without crossing any image of ab, it is clear that the whole infinite region 1′ should be considered as mapped on I′; and similarly 1 on I. The converse is also evident, for the same reason.

If, on reaching A, the point z turns to the left through 90° and moves along AC, then w will make a turn to the left of 180°, that is, will keep straight along ac; a turn as at R into 1′ will correspond to a turn as at r into I′. This checks with the statement that all 1′ is mapped on all I′. Suppose that z described a small circuit about $+1$. When z reaches D, w reaches d; when z reaches E, w reaches e. But when w crossed ac, it could not have crossed into I, and when it reaches e it cannot be in I; for the points of I are already accounted for as corresponding to points in 1. Hence in crossing ac, w must drop into one of the lower sheets, say the middle, II; and on reaching e it is still in II. It is thus seen that II corresponds to 2. Let z continue around its circuit; then II′ and 2′ correspond. When z crosses AC' from 2′ and moves into 1, the point w crosses ac' and moves from II′ up into I. In fact the upper two sheets are connected along ac just as the two sheets of the surface for $w = z^2$ were connected along their junction.

In like manner suppose that z moves from 0 to -1 and takes a turn about B so that w moves from 0 to 2 and takes a turn about b. When z crosses BF from 1′ to 3, w crosses bf from I′ into the upper half of some sheet, and this must be III for the reason that I and II are already mapped on 1 and 2. Hence I′ and III are connected, and so are I and III′. This leaves II which has been cut along bf, and III cut along ac, which may be reconnected as if they had never been cut. The reason for this appears forcibly if all the points z which correspond to the branch points are added to the diagram. When $w = 2$, the values of z are the critical value -1 (double) and the ordinary value $z = 2$; similarly, $w = -2$ corresponds to $z = -2$. Hence if z describe the half circuit AE so that w gets around to e in II, then if z moves out to $z = 2$, w will move out to $w = 2$, passing by $w = 0$ in the sheet II as z passes through $z = \sqrt{3}$; but as $z = 2$ is not a critical point, $w = 2$ in II cannot be a branch point, and the cut in II may be reconnected.

The w-surface thus constructed for $w = f(z) = z^3 - 3z$ is the Riemann surface for the inverse function $z = f^{-1}(w)$, of which the explicit form cannot be given without solving a cubic. To each point of the surface corresponds one value of z, and to the three superposed values of w correspond three different values of z except at the branch points where two of the sheets come together and give only one value of z while the third sheet gives one other. The Riemann surface could equally well have been constructed by joining the two branch points and then connecting one of them to ∞. The image of $w = 0$ would not have been changed. The connections of the sheets could be established as before, but would be different. If the junction line be $-2, 2, +\infty$, the point $w = 2$ has two junctions running into it, and the connections of the sheets on opposite sides of the point are not independent. It is advisable to arrange the work so that the first branch point

which is encircled shall have only one junction running from it. This may be done by taking a very large circuit in z so that w will describe a large circuit and hence cut only one junction line, namely, from 2 to ∞, or by taking a small circuit about $z = 1$ so that w will take a small turn about $w = -2$. Let the latter method be chosen. Let z start from $z = 0$ at O and move to $z = 1$ at A; then w starts at $w = 0$ and moves to $w = -2$. The correspondence between 1′ and I′ is thus established. Let z turn about A; then w turns about $w = -2$ at a. As the line -2 to $-\infty$ or ac is not now a junction line, w moves from I′ into the upper half I, and the region across AC from 1′ should be labeled 1 to correspond. Then 2′, 2 and II′, II may be filled in. The connections of I–II′ and II–I′ are indicated and III–III′ is reconnected, as the branch point is of the first order and only two sheets are involved. Now let z move from $z = 0$ to $z = -1$ and take a turn about B; then w moves from $w = 0$ to $w = 2$ and takes a turn about b. The region next 1′ is marked 3 and I′ is connected to III. Passing from 3 to 3′ for z is equivalent to passing from III to III′ for w between 0 and b where these sheets are connected. From 3′ into 2 for z indicates III′ to II across the junction from $w = 2$ to ∞. This leaves I and II′ to be connected across this junction. The connections are complete. They may be checked by allowing z to describe a large circuit so that the regions 1, 1′, 3, 3′, 2, 2′, 1 are successively traversed. That I, I′, III, III′, II, II′, I is the corresponding succession of sheets is clear from the connections between $w = 2$ and ∞ and the fact that from $w = -2$ to $-\infty$ there is no junction.



w–surface *z–plane*

Consider the function $w = z^6 - 3z^4 + 3z^2$. The critical points are $z = 0, 1, 1, -1, -1$ and the corresponding branch points are $w = 0, 1, 1, 1, 1$. Draw the junction lines from $w = 0$ to $-\infty$ and from $w = 1$ to $+\infty$ along the axis of reals. To find the image of $v = 0$ on the z-plane, polar coördinates may be used.

$$z = r(\cos\phi + i\sin\phi), \qquad w = u + iv = r^6e^{6\phi i} - 3r^4e^{4\phi i} + 3r^2e^{2\phi i}.$$

$$v = 0 = r^2[r^4\sin 6\phi - 3r^2\sin 4\phi + 3\sin 2\phi]$$
$$= r^2\sin 2\phi[r^4(3 - 4\sin 2\phi) - 6r^2\cos\phi + 3].$$

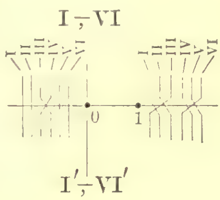
The equation $v = 0$ therefore breaks up into the equation $\sin 2\phi = 0$ and

$$r^2 = \frac{3\cos 2\phi \pm \sqrt{3}\sin 2\phi}{3 - 4\sin^2 2\phi} = \frac{\sqrt{3}}{2}\,\frac{\sin(60 \pm 2\phi)}{\sin(60 + 2\phi)\sin(60 - 2\phi)} = \frac{\sqrt{3}}{2\sin(60 \pm 2\phi)}.$$

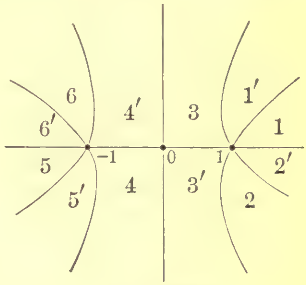
Hence the axes $\phi = 0°$ and $\phi = 90°$ and the two rectangular hyperbolas inclined at angles of $\pm 15°$ are the images of $v = 0$. The z-plane is thus divided into six portions. The function w is of the sixth order and six sheets must be spread over the w-plane and cut along the junction lines.

To connect up the sheets it is merely necessary to get a start. The line $w = 0$ to $w = 1$ is not a junction line and the sheets have not been cut through along it. But when z is small, real, and increasing, w is also small, real, and increasing. Hence to OA corresponds oa in any sheet desired. Moreover the region above OA will correspond to the upper half of the sheet and the region below OA to the lower half. Let the sheet be chosen as III and place the numbers 3 and 3′ so as to correspond with III and III′. Fill in the numbers 4 and 4′ around $z = 0$. When

z turns about the critical point $z = 0$, w turns about $w = 0$, but as angles are doubled it must go around twice and the connections III–IV′, IV–III′ must be made. Fill in more numbers about the critical point $z = 1$ of the second order where angles are tripled. On the w-surface there will be a triple connection III′–II, II′–I, I′–III. In like manner the critical point $z = -1$ may be treated. The surface is complete except for reconnecting sheets I, II, V, VI along $w = 0$ to $w = -\infty$ as if they had never been cut.



w—surface



z—plane

EXERCISES

1. Plot the corresponding lines for: $(\alpha)\ w = (1 + 2i)z$, $(\beta)\ w = (1 - \frac{1}{2}i)z$.

2. Solve for x and y in (1) and (2) of the text and plot the corresponding lines.

3. Plot the corresponding orthogonal systems of curves in these cases:

$$(\alpha)\ w = \frac{1}{z}, \qquad (\beta)\ w = 1 + z^2, \qquad (\gamma)\ w = \cos z.$$

4. Study the correspondence between z and w near the critical points:

$$(\alpha)\ w = z^3, \qquad (\beta)\ w = 1 - z^2, \qquad (\gamma)\ w = \sin z.$$

5. Upon the w-surface for $w = z^2$ plot the points corresponding to $z = 1,\ 1 + i,\ 2i,\ -\frac{1}{2} + \frac{1}{2}\sqrt{3}\,i,\ -\frac{1}{2},\ -\frac{1}{2}\sqrt{3} - \frac{1}{2}i,\ -i,\ \frac{1}{2} - \frac{1}{2}i$. And in the z-plane plot the points corresponding to $w = \sqrt{2} + \sqrt{2}\,i,\ i,\ -1,\ -\frac{1}{2} - \frac{1}{2}\sqrt{3}\,i,\ 1 - i$, whether in the upper or lower sheet.

6. Construct the w-surface for these functions:

$$(\alpha)\ w = z^3, \qquad (\beta)\ w = z^{-2}, \qquad (\gamma)\ w = 1 + z^2, \qquad (\delta)\ w = (z - 1)^3.$$

In (β) the singular point $z = 0$ should be joined by a cut to $z = \infty$.

7. Construct the Riemann surfaces for these functions:

$$(\alpha)\ w = z^4 - 2z^2, \qquad (\beta)\ w = -z^4 + 4z, \qquad (\gamma)\ w = 2z^5 - 5z^2,$$

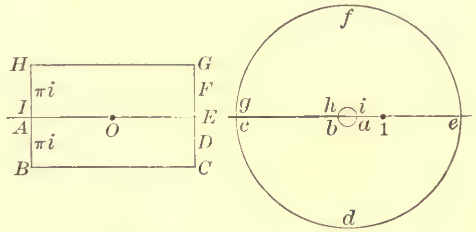
$$(\delta)\ w = z + \frac{1}{z}, \qquad (\epsilon)\ w = z^2 + \frac{1}{z^2}, \qquad (\zeta)\ w = \frac{z^3 + \sqrt{3}\,z}{\sqrt{3}\,z^2 + 1}.$$

185. Integrals and their inversion. Consider the function

$$z = \int_1^w \frac{dw}{w}, \qquad z = \ln w, \qquad w = \ln^{-1} z,$$

defined by an integral, and let the methods of the theory of functions be applied to the study of the function and its inverse. If w describes a path surrounding the origin, the integral need not vanish; for the

integrand is not analytic at $w = 0$. Let a cut be drawn from $w = 0$ to $w = -\infty$. The integral is then a single valued function of w provided the path of integration does not cross the cut. Moreover, it is analytic except at $w = 0$, where the derivative, which is the integrand $1/w$, ceases to be continuous. Let the w-plane as cut be mapped on the z-plane by allowing w to trace the path $1abcdefghi1$, by computing the value of z sufficiently to draw the image, and by applying the principles of conformal representation. When w starts from $w = 1$ and traces $1a$, z starts from $z = 0$ and becomes negatively very large. When w turns to the left to trace ab, z will turn also through 90° to the left. As the integrand along ab is $id\phi$, z must be changing by an amount which is pure imaginary and must reach B when w reaches b. When w traces bc, both w and dw are negative and z must be increasing by real positive quantities, that is, z must trace BC. When w moves along $cdefg$ the same reasoning as for the path ab will show that z moves along $CDEFG$. The remainder of the path may be completed by the reader.



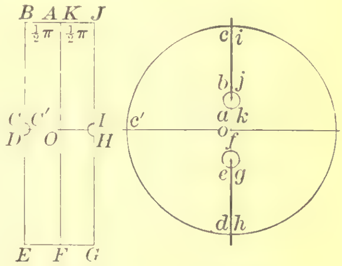
z—plane *w—plane*

It is now clear that the whole w-plane lying between the infinitesimal and infinite circles and bounded by the two edges of the cut is mapped on a strip of width $2\pi i$ bounded upon the right and left by two infinitely distant vertical lines. If w had made a complete turn in the positive direction about $w = 0$ and returned to its starting point, z would have received the increment $2\pi i$. That is to say, the values of z which correspond to the same point w reached by a direct path and by a path which makes k turns about $w = 0$ will differ by $2k\pi i$. Hence when w is regarded inversely as a function of z, the function will be periodic with the period $2\pi i$. It has been seen from the correspondence of $cdefg$ to $CDEFG$ that w becomes infinite when z moves off indefinitely to the right in the strip, and from the correspondence of $BAIH$ with $baih$ that w becomes 0 when z moves off to the left. Hence w must be a rational function of e^z. As w neither becomes infinite nor vanishes for any finite point of the strip, it must reduce merely to Ce^{kz} with k integral. As w has no smaller period than $2\pi i$, it follows that $k = 1$. To determine C, compare the derivative $dw/dz = Ce^z$ at $z = 0$ with its reciprocal $dz/dw = w^{-1}$ at the corresponding point $w = 1$; then $C = 1$. The inverse function $\ln^{-1}z$ is therefore completely determined as e^z.

In like manner consider the integral

$$z = \int_0^w \frac{dw}{1+w^2}, \qquad z = f(w), \qquad w = \phi(z) = f^{-1}(z).$$

Here the points $w = \pm i$ must be eliminated from the w-plane and the plane rendered simply connected by the proper cuts, say, as in the figure. The tracing of the figure may be left to the reader. The chief difficulty may be to show that the integrals along oa and bc are so nearly equal that C lies close to the real axis; no computation is really necessary inasmuch as the integral along oc' would be real and hence C' must lie on the axis. The image of the cut w-plane is a strip of width π. Circuits around either $+i$ or $-i$ add π to z, and hence w as a function of z has the period π. At the ends of the strip, w approaches the finite values $+i$ and $-i$. The function $w = \phi(z)$ has a simple zero when $z = 0$ and has no other zero in the strip. At the two points $z = \pm \frac{1}{2}\pi$, the function w becomes infinite, but only one of these points should be considered as in the strip. As the function has only one zero, the point $z = \frac{1}{2}\pi$ must be a pole of the first order. The function is therefore completely determined except for a constant factor which may be fixed by examining the derivative of the function at the origin. Thus



z–plane w–plane

$$w = c\frac{e^{2iz}-1}{e^{2iz}+1} = \frac{1}{i}\frac{e^{iz}-e^{-iz}}{e^{iz}+e^{-iz}} = \tan z, \qquad z = \tan^{-1} w.$$

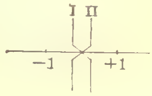
186. As a third example consider the integral

$$z = \int_0^w \frac{dw}{\sqrt{1-w^2}}, \qquad z = f(w), \qquad w = \phi(z) = f^{-1}(z). \qquad (8)$$

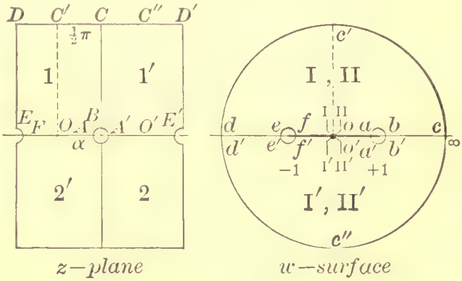
Here the integrand is double valued in w and consequently there is liable to be confusion of the two values in attempting to follow a path in the w-plane. Hence a two-leaved surface for the integrand will be constructed and the path of integration will be considered to be on the surface. Then to each point of the path there will correspond only one value of the integrand, although to each value of w there correspond two superimposed points in the two sheets of the surface.

As the radical $\sqrt{1-w^2}$ vanishes at $w = \pm 1$ and takes on only the single value 0 instead of two equal and opposite values, the points $w = \pm 1$ are branch points on the surface and they are the only finite branch points. Spread two sheets over the w-plane, mark the branch points $w = \pm 1$, and draw the junction line between them and continue it (provisionally) to $w = \infty$. At $w = -1$ the function $\sqrt{1-w^2}$ may be written $\sqrt{1+w}\,E(w)$, where E denotes a function which does not vanish at $w = -1$. Hence in the neighborhood of $w = -1$ the surface looks like that for $\sqrt{w}$ near $w = 0$. This may be accomplished by making the connections across the

junction line. At the point $w = +1$ the surface must cut through itself in a similar manner. This will be so provided that the sheets are reconnected across 1∞ as if never cut; if the sheets had been cross-connected along 1∞, each sheet would have been separate, though crossed, over 1, and the branch point would have disappeared. It is noteworthy that if w describes a large circuit including both branch points, the values of $\sqrt{1-w^2}$ are not interchanged; the circuit closes in each sheet without passing into the other. This could be expressed by saying that $w = \infty$ is not a branch point of the function.



Now let w trace out various paths on the surface in the attempt to map the surface on the z-plane by aid of the integral (8). To avoid any difficulties in the way of double or multiple values for z which might arise if w turned about a branch point $w = \pm 1$, let the surface be marked in each sheet over the axis of reals from $-\infty$ to $+1$. Let each of the four half planes be treated separately. Let w start at $w = 0$ in the upper half plane of the upper sheet and let the value of $\sqrt{1-w^2}$ at this point be $+1$; the values of $\sqrt{1-w^2}$ near $w = 0$ in II′ will then be near $+1$ and will be sharply distinguished from the values near -1 which are supposed to correspond to points in I′, II. As w traces oa, the integral z increases from 0 to a definite positive number α. The value of the integral from a to b is infinitesimal. Inasmuch as $w = 1$ is a branch point where two sheets connect, it is natural to assume that as w passes 1 and leaves it on the right, z will turn through half a straight angle. In other words the integral from b to c is naturally presumed to be a large pure imaginary affected with a positive sign. (This fact may easily be checked by examining the change in $\sqrt{1-w^2}$ when w describes a small circle about $w = 1$. In fact if the E-function $\sqrt{1+w}$ be discarded and if $1 - w$ be written as $re^{\phi i}$, then $\sqrt{r}e^{\frac{1}{2}\phi i}$ is that value of the radical which is positive when $1 - w$ is positive. Now when w describes the small semicircle, ϕ changes from 0° to −180° and hence the value of the radical along bc becomes $-i\sqrt{r}$ and the integrand is a positive pure imaginary.) Hence when w traces bc, z traces BC. At c there is a right-angle turn to the left, and as the value of the integral over the infinite quadrant cc' is $\frac{1}{2}\pi$, the point z will move back through the distance $\frac{1}{2}\pi$. That the point C' thus reached must lie on the pure imaginary axis is seen by noting that the integral taken directly along oc' would be pure imaginary. This shows that $\alpha = \frac{1}{2}\pi$ without any necessity of computing the integral over the interval oa. The rest of the map of I may be filled in at once by symmetry.

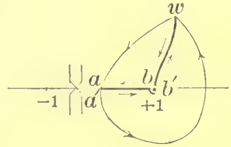


To map the rest of the w-surface is now relatively simple. For I′ let w trace $cc''d'$; then z will start at C and trace $CD' = \pi$. When w comes in along the lower side of the cut $d'e'$ in the upper sheet I′, the value of the integrand is identical with the value when this line de regarded as belonging to the upper half plane was described, for the line is not a junction line of the surface. The trace of z is therefore $D'E'$. When w traces $f'o'$ it must be remembered that I′ joins on to II and hence that the values of the integrand are the negative of those along fo. This

makes z describe the segment $F'O' = -\alpha = -\frac{1}{2}\pi$. The turn at $E'F'$ checks with the straight angle at the branch point -1. It is further noteworthy that when w returns to o' on I′, z does not return to 0 but takes the value π. This is no contradiction; the one-to-one correspondence which is being established by the integral is between points on the w-surface and points in a certain region of the z-plane, and as there are two points on the surface to each value of w, there will be two points z to each w. Thus far the sheet I has been mapped on the z-plane. To map II let the point w start at o' and drop into the lower sheet and then trace in this sheet the path which lies directly under the path it has traced in I. The integrand now takes on values which are the negatives of those it had previously, and the image on the z-plane is readily sketched in. The figure is self-explanatory. Thus the complete surface is mapped on a strip of width 2π.

To treat the different values which z may have for the same value of w, and in particular to determine the periods of w as the inverse function of z, it is necessary to study the value of the integral along different sorts of paths on the surface. Paths on the surface may be divided into two classes, closed paths and those not closed. A closed path is one which returns to the same point on the surface from which it started; it is not sufficient that it return to the same value of w. Of paths which are not closed on the surface, those which close in w, that is, which return to a point superimposed upon the starting point but in a different sheet, are the most important. These paths, on the particular surface here studied, may be further classified. A path which closes on the surface may either include neither branch point, or may include both branch points or may wind twice around one of the points. A path which closes in w but not on the surface may wind once about one of the branch points. Each of these types will be discussed.

If a closed path contains neither branch point, there is no danger of confusing the two values of the function, the projection of the path on the w-plane gives a region over which the integrand may be considered as single valued and analytic, and hence the value of the circuit integral is 0. If the path surrounds both branch points, there is again no danger of confusing the values of the function, but the projection of the path on the w-plane gives a region at two points of which, namely, the branch points, the integrand ceases to be analytic. The inference is that the value of the integral may not be zero and in fact will not be zero unless the integral around a circuit shrunk close up to the branch points or expanded out to infinity is zero. The integral around $cc'dc''c$ is here equal to 2π; the value of the integral around any path which incloses both branch points once and only once is therefore 2π or -2π according as the path lies in the upper or lower sheet; if the path surrounded the points k times, the value of the integral would be $2k\pi$. It thus appears that w regarded as a function of z has a period 2π. If a path closes in w but not on the surface, let the point where it crosses the junction line be held fast (figure) while the path is shrunk down to $wbaa'b'w$. The value of the integral will not change during this shrinking of the path, for the new and old paths may together be regarded as closed and of the first case considered. Along the paths wba and $a'b'w$ the integrand has opposite signs, but so has dw; around the small circuit the value of the integral is infinitesimal. Hence the value of the integral around the path which closes in w is $2I$ or $-2I$ if I is the value from the point a where the path crosses the junction line



to the point w. The same conclusion would follow if the path were considered to shrink down around the other branch point. Thus far the possibilities for z corresponding to any given w are $z+2k\pi$ and $2m\pi-z$. Suppose finally that a path turns twice around one of the branch points and closes on the surface. By shrinking the path, a new equivalent path is formed along which the integral cancels out term for term except for the small double circuit around ± 1 along which the value of the integral is infinitesimal. Hence the values $z+2k\pi$ and $2m\pi-z$ are the only values z can have for any given value of w if z be a particular possible value. This makes two and only two values of z in each strip for each value of w, and the function is of the second order.

It thus appears that w, as a function of z, has the period 2π, is single valued, becomes infinite at both ends of the strip, has no singularities within the strip, and has two simple zeros at $z=0$ and $z=\pi$. Hence w is a rational function of e^{iz} with the numerator $e^{2iz}-1$ and the denominator $e^{2iz}+1$. In fact

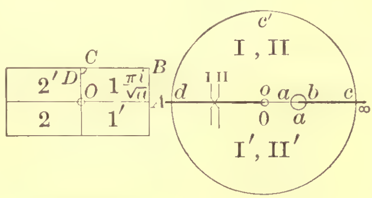
$$w=C\frac{e^{iz}-e^{-iz}}{e^{iz}+e^{-iz}}=\frac{1}{i}\frac{e^{iz}-e^{-iz}}{e^{iz}+e^{-iz}}=\sin z.$$

The function, as in the previous cases, has been wholly determined by the general methods of the theory of functions without even computing α.

One more function will be studied in brief. Let

$$z=\int_0^w\frac{dw}{(a-w)\sqrt{w}},\qquad a>0,\qquad z=f(w),\qquad w=\phi(z)=f^{-1}(z).$$

Here the Riemann surface has a branch point at $w=0$ and in addition there is the singular point $w=a$ of the integrand which must be cut out of both sheets. Let the surface be drawn with a junction line from $w=0$ to $w=-\infty$ and with a cut in each sheet from $w=a$ to $w=\infty$. The map on the z-plane now becomes as indicated in the figure. The different values of z for the same value of w are readily seen to arise when w turns about the point $w=a$ in either sheet or when a path closes in w but not on the surface. These values of z are $z+2k\pi i/\sqrt{a}$ and $2m\pi i/\sqrt{a}-z$. Hence w as a function of z has the period $2\pi i a^{-\frac{1}{2}}$, has a zero at $z=0$ and a pole at $z=\pi i/\sqrt{a}$, and approaches the finite value $w=a$ at both ends of the strip. It must be noted, however, that the zero and pole are both necessarily double, for to any ordinary value of w correspond two values of z in the strip. The function is therefore again of the second order, and indeed



z—plane *w—surface*

$$w=a\frac{(e^{z\sqrt{a}}-1)^2}{(e^{z\sqrt{a}}+1)^2}=a\tanh^2\frac{1}{2}z\sqrt{a},\qquad z=\frac{2}{\sqrt{a}}\tanh^{-1}\sqrt{\frac{w}{a}}.$$

The success of this method of determining the function $z=f(w)$ defined by an integral, or the inverse $w=f^{-1}(z)=\phi(z)$, has been dependent first upon the ease with which the integral may be used to map the w-plane or w-surface upon the z-plane, and second upon the simplicity of the map, which was such as to indicate that the inverse function was a single valued periodic function. It should be

realized that if an attempt were made to apply the methods to integrands which appear equally simple, say to

$$z = \int \sqrt{a^2 - w^2}\,dw, \qquad z = \int (a - w)\,dw/\sqrt{w},$$

the method would lead only with great difficulty, if at all, to the relation between z and w; for the functional relation between z and w is indeed not simple. There is, however, one class of integrals of great importance, namely,

$$z = \int \frac{dw}{\sqrt{(w - \alpha_1)(w - \alpha_2) \cdots (w - \alpha_n)}}$$

for which this treatment is suggestive and useful.

EXERCISES

1. Discuss by the method of the theory of functions these integrals and inverses:

$(\alpha)\ \int_1^w \frac{dw}{2w}$. $(\beta)\ \int_0^w \frac{2\,dw}{1 - w}$, $(\gamma)\ \int_0^w \frac{dw}{1 - w^2}$.

$(\delta)\ \int_0^w \frac{dw}{\sqrt{w^2 - 1}}$. $(\epsilon)\ \int_0^w \frac{dw}{\sqrt{w^2 + 1}}$. $(\zeta)\ \int_\infty^w \frac{dw}{w\sqrt{w^2 + a^2}}$,

$(\eta)\ \int_\infty^w \frac{dw}{w\sqrt{w^2 - a^2}}$. $(\theta)\ \int_0^w \frac{dw}{\sqrt{2aw - w^2}}$. $(\iota)\ \int_1^w \frac{dw}{(w + 1)\sqrt{w^2 - 1}}$.

The results may be checked in each case by actual integration.

2. Discuss $\int_\infty^w \frac{dw}{\sqrt{w(1 - w)(1 + w)}}$ and $\int_0^w \frac{dw}{\sqrt{1 - w^4}}$ (§ 182, and Ex. 10, p. 489).

CHAPTER XIX

ELLIPTIC FUNCTIONS AND INTEGRALS

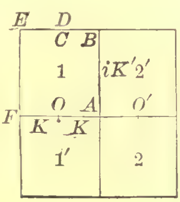
187. Legendre's integral I and its inversion. Consider

$$z = \int_0^w \frac{dw}{\sqrt{(1 - w^2)(1 - k^2w^2)}}, \qquad 0 < k < 1. \tag{I}$$

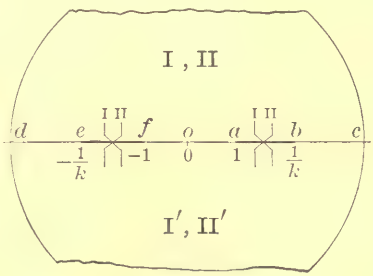
The Riemann surface for the integrand* has branch points at $w = \pm 1$ and $\pm 1/k$ and is of two sheets. Junction lines may be drawn between $+1$, $+1/k$ and -1, $-1/k$. For very large values of w, the radical $\sqrt{(1 - w^2)(1 - k^2w^2)}$ is approximately $\pm kw^2$ and hence there is no danger of confusing the values of the function. Across the junction lines the surface may be connected as indicated, so that in the neighborhood of $w = \pm 1$ and $w = \pm 1/k$ it looks like the surface for $\sqrt{w}$. Let $+1$ be the value of the integrand at $w = 0$ in the upper sheet. Further let

$$K = \int_0^1 \frac{dw}{\sqrt{(1 - w^2)(1 - k^2w^2)}}, \quad iK' = \int_1^{\frac{1}{k}} \frac{dw}{\sqrt{(1 - w^2)(1 - k^2w^2)}}. \tag{1}$$

Let the changes of the integral be followed so as to map the surface on the z-plane. As w moves from o to a, the integral (I) increases by K, and z moves from O to A. As w continues straight on, z makes a right-angle turn and increases by pure imaginary increments to the total amount iK' when w reaches b. As w continues there is



z—*plane*



w—*surface*

another right-angle turn in z, the integrand again becomes real, and z moves down to C. (That z reaches C follows from the facts that the

* The reader unfamiliar with Riemann surfaces (§ 184) may proceed at once to identify (I) and (2) by Ex. 9, p. 475 and may take (1) and other necessary statements for granted.

integral along an infinite quadrant is infinitesimal and that the direct integral from 0 to $i\infty$ would be pure imaginary like dw.) If w is allowed to continue, it is clear that the map of I will be a rectangle $2K$ by K' on the z-plane. The image of all four half planes of the surface is as indicated. The conclusion is reasonably apparent that w as the inverse function of z is doubly periodic with periods $4K$ and $2iK'$.

The periodicity may be examined more carefully by considering different possibilities for paths upon the surface. A path surrounding the pairs of branch points 1 and k^{-1} or -1 and $-k^{-1}$ will close on the surface, but as the integrand has opposite signs on opposite sides of the junction lines, the value of the integral is $2iK'$. A path surrounding -1, $+1$ will also close; the small circuit integrals about -1 or $+1$ vanish and the integral along the whole path, in view of the opposite values of the integrand along fa in I and II, is twice the integral from f to a or is $4K$. Any path which closes on the surface may be resolved into certain multiples of these paths. In addition to paths which close on the surface, paths which close in w may be considered. Such paths may be resolved into those already mentioned and paths running directly between 0 and w in the two sheets. All possible values of z for any w are therefore $4mK + 2niK' \pm z$. The function $w(z)$ has the periods $4K$ and $2iK'$, is an odd function of z as $w(-z) = w(z)$, and is of the second order. The details of the discussion of various paths is left to the reader.

Let $w = f(z)$. The function $f(z)$ vanishes, as may be seen by the map, at the two points $z = 0$, $2K$ of the rectangle of periods, and at no other points. These zeros of w are simple, as $f'(z)$ does not vanish. The function is therefore of the second order. There are poles at $z = iK', 2K + iK'$, which must be simple poles. Finally $f(K) = 1$. The position of the zeros and poles determines the function except for a constant multiplier, and that will be fixed by $f(K) = 1$; the function is wholly determined. The function $f(z)$ may now be identified with $\operatorname{sn} z$ of § 177 and in particular with the special case for which K and K' are so related that the multiplier $g = 1$.

$$w = f(z) = \frac{\Theta(K)}{H(K)}\frac{H(z)}{\Theta(z)} = \operatorname{sn} z, \qquad z = u. \tag{2}$$

For the quotient of the theta functions has simple zeros at 0, $2K$, where the numerator vanishes, and simple poles at iK', $2K + iK'$, where the denominator vanishes; the quotient is 1 at $z = K$; and the derivative of $\operatorname{sn} z$ at $z = 0$ is $g \operatorname{cn} 0 \operatorname{dn} 0 = g = 1$, whereas $f'(0) = 1$ is also 1.

The imposition of the condition $g = 1$ was seen to impose a relation between K, K', k, k', g by virtue of which only one of the five remained independent. The definition of K and K' as definite integrals also makes them functions $K(k)$ and $K'(k)$ of k. But

$$iK'(k) = \int_1^{\frac{1}{k}} \frac{dw}{\sqrt{(1-w^2)(1-k^2w^2)}} \qquad\qquad (3)$$
$$= i\int_0^1 \frac{dw_1}{\sqrt{(1-w_1^2)(1-k'^2w_1^2)}} = iK(k')$$

if $w = (1-k'^2w_1^2)^{\frac{1}{2}}$ and $k^2 + k'^2 = 1$. Hence it appears that K may be computed from k' as K' from k. This is very useful in practice when k^2 is near 1 and k'^2 near 0. Thus let

$$e^{-\pi\frac{K}{K'}} = q' = \frac{1}{2}\frac{1-\sqrt{k}}{1+\sqrt{k}} + \frac{2}{2^5}\left(\frac{1-\sqrt{k}}{1+\sqrt{k}}\right)^5 + \cdots, \qquad \log q \log q' = \pi^2, \qquad (4)$$
$$\sqrt{\frac{2K'}{\pi}} = \Theta_1(0, q') = 1 + 2q' + 2q'^4 + \cdots, \qquad K = -\frac{K'}{\pi}\log q';$$

and compare with (37) of p. 472. Now either k or k' is greater than 0.7, and hence either q or q' may be obtained to five places with only one term in its expansion and with a relative error of only about 0.01 per cent. Moreover either q or q' will be less than 1/20 and hence a single term $1 + 2q$ or $1 + 2q'$ gives K or K' to four places.

188. As in the relation between the Riemann surface and the z-plane the whole real axis of z corresponds periodically to the part of the real axis of w between -1 and $+1$, the function $\operatorname{sn} x$, for real x, is real. The graph of $y = \operatorname{sn} x$ has roots at $x = 2mK$, maxima or minima alternately at $(2m+1)K$, inflections inclined at the angle 45° at the roots, and in general looks like $y = \sin(\pi x/2K)$. Examined more closely, $\operatorname{sn}\frac{1}{2}K = (1+k')^{-\frac{1}{2}} > 2^{-\frac{1}{2}} = \sin\frac{1}{4}\pi$; it is seen that the curve $\operatorname{sn} x$ has ordinates numerically greater than $\sin(\pi x/2K)$. As

$$\operatorname{cn} x = \sqrt{1-\operatorname{sn}^2 x}, \qquad \operatorname{dn} x = \sqrt{1-k^2\operatorname{sn}^2 x}, \qquad (5)$$

the curves $y = \operatorname{cn} x$, $y = \operatorname{dn} x$, may readily be sketched in. It may be noted that as $\operatorname{sn}(x+K) \neq \operatorname{cn} x$, the curves for $\operatorname{sn} x$ and $\operatorname{cn} x$ cannot be superposed as in the case of the trigonometric functions.

The segment 0, iK' of the pure imaginary axis for z corresponds to the whole upper half of the pure imaginary axis for w. Hence $\operatorname{sn} ix$ with x real is pure imaginary and $-i\operatorname{sn} ix$ is real and positive for $0 \leqq x < K'$ and becomes infinite for $x = K'$. Hence $-i\operatorname{sn} ix$ looks in general like $\tan(\pi x/2K')$. By (5) it is seen that the curves for $y = \operatorname{cn} ix$, $y = \operatorname{dn} ix$ look much like $\sec(\pi x/2K')$ and that $\operatorname{cn} ix$ lies above $\operatorname{dn} ix$. These functions are real for pure imaginary values.

It was seen that when k and k' interchanged, K and K' also interchanged. It is therefore natural to look for a relation between the elliptic functions $\operatorname{sn}(z, k)$, $\operatorname{cn}(z, k)$, $\operatorname{dn}(z, k)$ formed with the modulus k

and the functions $\operatorname{sn}(z, k')$, $\operatorname{cn}(z, k')$, $\operatorname{dn}(z, k')$ formed with the complementary modulus k' It will be shown that

$$\operatorname{sn}(iz, k) = i\frac{\operatorname{sn}(z, k')}{\operatorname{cn}(z, k')}, \qquad \operatorname{sn}(z, k) = -i\frac{\operatorname{sn}(iz, k')}{\operatorname{cn}(iz, k')},$$

$$\operatorname{cn}(iz, k) = \frac{1}{\operatorname{cn}(z, k')}, \qquad \operatorname{cn}(z, k) = \frac{1}{\operatorname{cn}(iz, k')},$$

$$\operatorname{dn}(iz, k) = \frac{\operatorname{dn}(z, k')}{\operatorname{cn}(z, k')}, \qquad \operatorname{dn}(z, k) = \frac{\operatorname{dn}(iz, k')}{\operatorname{cn}(iz, k')}.$$

Consider $\operatorname{sn}(iz, k)$. This function is periodic with the periods $4K$ and $2iK'$ if iz be the variable, and hence with periods $4iK$ and $2K'$ if z be the variable. With z as variable it has zeros at 0, $2iK$, and poles at K', $2iK + K'$. These are precisely the positions of the zeros and poles of the quotient $H(z, q')/H_1(z, q')$, where the theta functions are constructed with q' instead of q. As this quotient and $\operatorname{sn}(iz, k)$ are of the second order and have the same periods,

$$\operatorname{sn}(iz, k) = C\frac{H(z, q')}{H_1(z, q')} = C_1\frac{\operatorname{sn}(z, k')}{\operatorname{cn}(z, k')}.$$

The constant C_1 may be determined as $C_1 = i$ by comparing the derivatives of the two sides at $z = 0$. The other five relations may be proved in the same way or by transformation.

The theta series converge with extreme rapidity if q is tolerably small, but if q is somewhat larger, they converge rather poorly. The relations just obtained allow the series with q to be replaced by series with q' and one of these quantities is surely less than $1/20$. In fact if $\nu = \pi x/2K$ and $\nu' = \pi x/2K'$, then

$$\begin{aligned}\operatorname{sn}(x, k) &= \frac{\sqrt[4]{q}}{\sqrt{k}}\,\frac{2\sin\nu - 2q^2\sin 3\nu + 2q^6\sin 5\nu - \cdots}{1 - 2q\cos 2\nu + 2q^4\cos 4\nu - 2q^9\cos 6\nu + \cdots} \\ &= \frac{1}{\sqrt{k}}\,\frac{\sinh\nu' - q'^2\sinh 3\nu' + q'^6\sinh 5\nu' - \cdots}{\cosh\nu' + q'^2\cosh 3\nu' + q'^6\cosh 5\nu' + \cdots}.\end{aligned} \tag{6}$$

The second series has the disadvantage that the hyperbolic functions increase rapidly, and hence if the convergence is to be as good as for the first series, the value of q' must be considerably less than that of q, that is, K' must be considerably less than K. This can readily be arranged for work to four or five places. For

$$q'^6 = e^{-6\pi\frac{K}{K'}}, \qquad \cosh 5\nu' = \tfrac{1}{2}\left(e^{\frac{5\pi x}{2K'}} + e^{-\frac{5\pi x}{2K'}}\right), \qquad 0 \leqq x \leqq K',$$

where owing to the periodicity of the functions it is never necessary to take $x > K'$. The term in q'^6 is therefore less than $\frac{1}{2}q'^{3\frac{1}{2}}$. If the term

in q'^6 is to be equally negligible with that in q^6,

$$2q^6 = \tfrac{1}{2}q'^{\frac{7}{2}} \quad \text{with} \quad \log q \log q' = \pi^2,$$

from which q' is determined as about $q' = .02$ and q as about $q = .08$; the neglected term is about 0.0000005 and is barely enough to effect six-place work except through the multiplication of errors. The value of k corresponding to this critical value of q is about $k = 0.85$.

Another form of the integral under consideration is

$$F(\phi, k) = \int_0^\phi \frac{d\theta}{\sqrt{1 - k^2 \sin^2\theta}} = \int_0^y \frac{dv}{\sqrt{1 - v^2}\sqrt{1 - k^2v^2}} = x, \quad (7)$$

$$\sin\phi = y = \operatorname{sn} x, \qquad \phi = \operatorname{am} x, \qquad \cos\phi = \sqrt{1 - \operatorname{sn}^2 x} = \operatorname{cn} x,$$

$$\Delta\phi = \sqrt{1 - k^2y^2} = \sqrt{1 - k^2\sin^2\phi} = \operatorname{dn} x, \qquad k'^2 = 1 - k^2,$$

$$x = \operatorname{sn}^{-1}(y, k) = \operatorname{cn}^{-1}(\sqrt{1 - y^2}, k) = \operatorname{dn}^{-1}(\sqrt{1 - k^2y^2}, k).$$

The angle ϕ is called the *amplitude* of x; the functions $\operatorname{sn} x$, $\operatorname{cn} x$, $\operatorname{dn} x$ are the *sine-amplitude, cosine-amplitude, delta-amplitude* of x. The half periods are then

$$K = \int_0^{\frac{1}{2}\pi} \frac{d\theta}{\sqrt{1 - k^2\sin^2\theta}} = F\left(\frac{1}{2}\pi, k\right),$$
$$K' = \int_0^{\frac{1}{2}\pi} \frac{d\theta}{\sqrt{1 - k'^2\sin^2\theta}} = F\left(\frac{1}{2}\pi, k'\right), \quad (8)$$

and are known as the *complete elliptic integrals of the first kind.*

189. The elliptic functions and integrals often arise in problems that call for a numerical answer. Here k^2 is given and the complete integral K or the value of the elliptic functions or of the elliptic integral $F(\phi, k)$ are desired for some assigned argument. The values of K and $F(\phi, k)$ in terms of $\sin^{-1}k$ are found in tables (B. O. Peirce, pp. 117–119), and may be obtained therefrom. The tables may be used by inversion to find the values of the function $\operatorname{sn} x$, $\operatorname{cn} x$, $\operatorname{dn} x$ when x is given: for $\operatorname{sn} x = \operatorname{sn} F(\phi, k) = \sin\phi$, and if $x = F$ is given, ϕ may be found in the table, and then $\operatorname{sn} x = \sin\phi$. It is, however, easy to compute the desired values directly, owing to the extreme rapidity of the convergence of the series. Thus

$$\sqrt{\frac{2K}{\pi}} = \Theta_1(0), \quad \sqrt{\frac{2Kk'}{\pi}} = \Theta(0), \quad \frac{1 + \sqrt{k'}}{\sqrt{2\pi}}\sqrt{K} = \frac{1}{2}(\Theta_1(0) + \Theta(0)),$$

$$\sqrt{K} = \frac{\sqrt{2\pi}}{1 + \sqrt{k'}}(1 + 2q^4 + \cdots) = \sqrt{-\frac{K'}{\pi}\log q'} \quad (9)$$
$$= \frac{\sqrt{-2\log q'}}{1 + \sqrt{k}}(1 + 2q'^4 + \cdots).$$

The elliptic functions are computed from (6) or analogous series. To compute the value of the elliptic integral $F(\phi, k)$, note that if

$$\cot\lambda = \frac{\operatorname{dn} x}{\sqrt{k'}} = \frac{1 + 2q\cos 2\nu + 2q^4\cos 4\nu + \cdots}{1 - 2q\cos 2\nu + 2q^4\cos 4\nu + \cdots}, \tag{10}$$

$$\tan\left(\frac{1}{4}\pi - \lambda\right) = \frac{\cot\lambda - 1}{\cot\lambda + 1} = 2q\,\frac{\cos 2\nu + q^8\cos 6\nu + \cdots}{1 + 2q^4\cos 4\nu + \cdots};$$

and $\tan(\frac{1}{4}\pi - \lambda) = 2q\cos 2\nu$ or $\tan(\frac{1}{4}\pi - \lambda) = \dfrac{2q\cos 2\nu}{1 + 2q^4\cos 4\nu}$ (10′)

are two approximate equations from which $\cos 2\nu$ may be obtained; the first neglects q^4 and is generally sufficient, but the second neglects only q^8. If k^2 is near 1, the proper approximations are

$$\cot\lambda = \frac{1}{\sqrt{k}}\frac{\operatorname{dn}(x, k)}{\operatorname{cn}(x, k)} = \frac{\operatorname{dn}(ix, k')}{\sqrt{k}} = \frac{1 + 2q'\cosh 2\nu' + \cdots}{1 - 2q'\cosh 2\nu' + \cdots}, \tag{11}$$

$$\tan(\tfrac{1}{4}\pi - \lambda) = 2q'\cosh 2\nu' \quad\text{or}\quad \tan(\tfrac{1}{4}\pi - \lambda) = \frac{2q'\cosh 2\nu'}{1 + 2q'^4\cosh 4\nu'}. \tag{11′}$$

Here $q'^8\cosh 8\nu' < q'^4$ is neglected in the second, but $q'^4\cosh 4\nu' < q'^2$ in the first, which is not always sufficient for four-place work. Of course if ϕ with $\operatorname{sn} x = \sin\phi$ or if $y = \operatorname{sn} x$ is given, $\operatorname{dn} x = \sqrt{1 - k^2\operatorname{sn}^2 x}$ and $\operatorname{cn} x = \sqrt{1 - \operatorname{sn}^2 x}$ are readily computed.

As an example take $\displaystyle\int_0^\phi \frac{d\theta}{\sqrt{1 - \frac{1}{4}\sin^2\theta}}$ and find K, $\operatorname{sn}\frac{2}{3}K$, $F(\frac{1}{8}\pi, \frac{1}{2})$. As $k'^2 = \frac{3}{4}$ and $\sqrt{k'} > 0.9$, the first term of (37), p. 472, gives q accurately to five places. Compute in the form: ($\mathrm{Lg} = \log_{10}$)

$\mathrm{Lg}\, k'^2 = 9.87506$	$\mathrm{Lg}\,(1 - \sqrt{k'}) = 8.84136$	$\mathrm{Lg}\, 2\pi = 0.7982$
$\mathrm{Lg}\sqrt{k'} = 9.96876$	$\mathrm{Lg}\,(1 + \sqrt{k'}) = 0.28569$	$2\,\mathrm{Lg}\,(1 + \sqrt{k'}) = 0.5714$
$\sqrt{k'} = 9.93060$	$\mathrm{Lg}\, 2q = 8.55567$	$\mathrm{Lg}\, K = 0.2268$
$1 - \sqrt{k'} = 0.06940$	$2q = 0.03595$	$K = 1.686$
$1 + \sqrt{k'} = 1.93060$	$q = 0.01797$	Check with table.

$$\operatorname{sn}\frac{2}{3}K = 2\,\frac{\sqrt[4]{q}}{\sqrt{k}}\,\frac{\sin\frac{1}{3}\pi - q^2\sin\pi + \cdots}{1 - 2q\cos\frac{2}{3}\pi + \cdots} = 2\cdot\frac{\sqrt[4]{q}}{\sqrt{\frac{1}{2}}}\,\frac{\frac{1}{2}\sqrt{3}}{1 + q}.$$

$\operatorname{sn}\dfrac{2}{3}K = \dfrac{\sqrt{6}\,\sqrt[4]{q}}{1.01797}$	$\frac{1}{2}\,\mathrm{Lg}\, 6 = 0.38908$	$\mathrm{Lg}\,\operatorname{sn}\frac{2}{3}K = 9.9450$
	$\frac{1}{4}\,\mathrm{Lg}\, q = 9.56366$	$\operatorname{sn}\frac{2}{3}K = 0.8810.$
	$-\,\mathrm{Lg}\, 1.018 = 9.99226$	

$\phi = \frac{1}{8}\pi \qquad \Delta\phi = \operatorname{dn} x = \sqrt{1 - \frac{1}{4}\sin^2\frac{1}{8}\pi} = \sqrt{1 - \frac{1}{2}\sin\frac{1}{8}\pi}\,\sqrt{1 + \frac{1}{2}\sin\frac{1}{8}\pi}.$

$\frac{1}{2}\sin\frac{1}{8}\pi = 0.19134$	$\lambda = 43^\circ\, 28'\, 28''$	$\mathrm{Lg}\, 42.20 = 1.6253$
$1 - \frac{1}{2}\sin\frac{1}{8}\pi = 0.80866$	$\frac{1}{4}\pi - \lambda = 1^\circ\, 31'\, 32''$	$\mathrm{Lg}\, K = 0.2268$
$1 + \frac{1}{2}\sin\frac{1}{8}\pi = 1.19134$	$\mathrm{Lg}\tan = 8.42540$	$-\mathrm{Lg}\, 180 = 7.7447$
$\frac{1}{2}\mathrm{Lg}\,(1 - \frac{1}{2}\sin\frac{1}{8}\pi) = 9.95388$	$\mathrm{Lg}\, 2\, q = 8.55567$	$\mathrm{Lg}\, x = 9.5968$
$\frac{1}{2}\mathrm{Lg}\,(1 + \frac{1}{2}\sin\frac{1}{8}\pi) = 0.03802$	$\mathrm{Lg}\cos 2\,\nu = 9.86973$	$x = 0.3952$
$-\mathrm{Lg}\sqrt{k'} = 0.03124$	$2\,\nu = 42^\circ\, 12'$	Check with table.
$\mathrm{Lg}\cot\lambda = 0.02314$	$180\, x = K\,(42.20)$	

As a second example consider a pendulum of length a oscillating through an arc of 300°. Find the period, the time when the pendulum is horizontal, and its position after dropping for a third of the time required for the whole descent. Let $x^2 + y^2 = 2\, ay$ be the equation of the path and $h = a\left(1 + \frac{1}{2}\sqrt{3}\right)$ the greatest height. When $y = h$, the energy is wholly potential and equals mgh; and mgy is the general value of the potential energy. The kinetic energy is

$$\frac{m}{2}\left(\frac{ds}{dt}\right)^2 = \frac{\frac{1}{2}\, ma^2}{2\, ay - y^2}\left(\frac{dy}{dt}\right)^2 \quad \text{and} \quad \frac{\frac{1}{2}\, ma^2}{2\, ay - y^2}\left(\frac{dy}{dt}\right)^2 + mgy = mgh$$

is the equation of motion by the principle of energy. Hence

$$t = \int_0^y \frac{a\,dy}{\sqrt{2\,g}\sqrt{(h-y)\,(2\,ay - y^2)}} = \sqrt{\frac{a}{g}}\int_0^w \frac{dw}{\sqrt{(1-w^2)\,(1-k^2w^2)}}, \quad w^2 = \frac{y}{h}, \quad k^2 = \frac{h}{2\,a},$$

$$\sqrt{g/a}\,t = \mathrm{sn}^{-1}(w, k), \qquad w = \mathrm{sn}\left(\sqrt{g/a}\,t, k\right), \qquad y = h\,\mathrm{sn}^2\left(\sqrt{g/a}\,t, k\right),$$

are the integrated results. The quarter period, from highest to lowest point, is $K\sqrt{a/g}$; the horizontal position is $y = a$, at which t is desired; and the position for $\sqrt{g/a}\,t = \frac{2}{3}K$ is the third thing required.

$$k^2 = 0.93301, \qquad 2\, q' = \frac{1 - \sqrt{k}}{1 + \sqrt{k}}, \qquad K = -\frac{K'}{\pi}\log q' = \frac{-2\,\mathrm{Lg}\, q'}{M\left(1 + \sqrt{k}\right)^2}.$$

$\mathrm{Lg}\, k^2 = 9.96988$	$\mathrm{Lg}\left(1 - \sqrt{k}\right) = 8.23553$	$\mathrm{Lg}\, 2 = 0.3010$
$\mathrm{Lg}\sqrt{k} = 9.99247$	$-\mathrm{Lg}\left(1 + \sqrt{k}\right) = 9.70272$	$\mathrm{Lg}^2\, q'^{-1} = 0.3734$
$\sqrt{k} = 0.98280$	$-\mathrm{Lg}\, 2 = 9.69897$	$-\mathrm{Lg}\, M = 0.3622$
$1 - \sqrt{k} = 0.01720$	$\mathrm{Lg}\, q' = 7.63722$	$-2\,\mathrm{Lg}\left(1 + \sqrt{k}\right) = 9.4034$
$1 + \sqrt{k} = 1.98280$	$q' = 0.00434$	$\mathrm{Lg}\, K = 0.4420.$

Hence $K = 2.768$ and the complete periodic time is $4\, K\sqrt{a/g}$.

$$y = a, \qquad w^2 = \frac{a}{h}, \qquad \mathrm{cn}\, w = \sqrt{1 - a/h}, \qquad \mathrm{dn}\, w = \sqrt{1 - k^2 a/h}.$$

$$\frac{1}{\sqrt{k}}\frac{\mathrm{dn}\, w}{\mathrm{cn}\, w} = \sqrt[4]{\frac{4}{3}k^2} = \cot\lambda, \qquad \tan\left(\frac{1}{4}\pi - \lambda\right) = 2\, q'\cosh 2\,\nu', \qquad 2\,\nu' = \frac{\pi K}{K'}\sqrt{\frac{g}{a}}\frac{t}{K}.$$

$\mathrm{Lg}\, k^2 = 9.96988$	$\lambda = 43^\circ\, 26'\, 12''$	$2\,\nu' = 1.813$
$\mathrm{Lg}\, 4 = 0.60206$	$\frac{1}{4}\pi - \lambda = 1^\circ\, 33'\, 48''$	$\mathrm{Lg}\, 2\,\nu' = 0.2584$
$-\mathrm{Lg}\, 3 = 9.52288$	$\mathrm{Lg}\tan = 8.43603$	$-\mathrm{Lg}^2\, q'^{-1} = 9.6266$
$\mathrm{Lg}\cot^4\lambda = 0.09482$	$\mathrm{Lg}\, 2\, q' = 9.93825$	$\mathrm{Lg}\, M = 9.6378$
$\mathrm{Lg}\cot\lambda = 0.02370$	$\mathrm{Lg}\cosh 2\,\nu' = 0.49778$	$\mathrm{Lg}\sqrt{\frac{g}{a}}\frac{t}{K} = 9.5228.$

Hence the time for $y = a$ is $t = 0.3333\, K\sqrt{a/g} = \frac{1}{3}$ whole time of ascent.

$$y = h\,\mathrm{sn}^2\sqrt{\frac{g}{a}}\,\frac{2}{3}K\sqrt{\frac{a}{g}} = \frac{h}{k}\left(\frac{\sinh \pi K/3K' - q'^2 \sinh \pi K/K'}{\cosh \pi K/3K' + q'^2 \cosh \pi K/K'}\right)^2$$

$$= 2ak\left(\frac{q'^{-\frac{1}{3}} - q'^{\frac{1}{3}} - q'^2(q'^{-1} - q')}{q'^{-\frac{1}{3}} + q'^{\frac{1}{3}} + q'^2(q'^{-1} + q')}\right)^2 = 2ak\left(\frac{q'^{-\frac{1}{3}} - q'^{\frac{1}{3}} - q'}{q'^{-\frac{1}{3}} + q'^{\frac{1}{3}} + q'}\right)^2.$$

$$\tfrac{1}{3}\,\mathrm{Lg}\,q' = 9.21241 \qquad q'^{\frac{1}{3}} = 0.1631 \qquad\qquad y = 2ak\left(\frac{5.9645}{6.2993}\right)^2.$$
$$-\tfrac{1}{3}\,\mathrm{Lg}\,q' = 0.78759 \qquad q'^{-\frac{1}{3}} = 6.1319$$

This gives $y = 1.732\,a$, which is very near the top at $h = 1.866\,a$. In fact starting at 30° from the vertical the pendulum reaches 43° in a third and 90° in another third of the total time of descent. As $\mathrm{sn}\,\frac{1}{2}K$ is $(1 + k')^{-\frac{1}{2}}$ it is easy to calculate the position of the pendulum at half the total time of descent.

EXERCISES

1. Discuss these integrals by the method of mapping:

(α) $z = \int_0^w \frac{dw}{\sqrt{(a^2 - w^2)(b^2 - w^2)}}$, $\quad a > b > 0, \quad w = b\,\mathrm{sn}\,az, \quad k = \frac{b}{a}$.

(β) $z = \int_0^w \frac{dw}{\sqrt{w(1 - w)(1 - k^2 w)}}$, $\quad w = \mathrm{sn}^2\left(\frac{1}{2}z, k\right), \quad z = 2\,\mathrm{sn}^{-1}(\sqrt{x}, k)$.

(γ) $z = \int_0^w \frac{dw}{\sqrt{(1 + w^2)(1 + k'^2 w^2)}}$, $\quad w = \frac{\mathrm{sn}\,(z, k)}{\mathrm{cn}\,(z, k)} = \mathrm{tn}\,(z, k), \quad z = \mathrm{tn}^{-1}(w, k)$.

2. Establish these Maclaurin developments with the aid of § 177:

(α) $\mathrm{sn}\,z = z - (1 + k^2)\frac{z^3}{3!} + (1 + 14k^2 + k^4)\frac{z^5}{5!} - \cdots$.

(β) $\mathrm{cn}\,z = 1 - \frac{z^2}{2!} + (1 + 4k^2)\frac{z^4}{4!} - (1 + 44k^2 + 16k^4)\frac{z^6}{6!} + \cdots$,

(γ) $\mathrm{dn}\,z = 1 - k^2\frac{z^2}{2!} + k^2(4 + k^2)\frac{z^4}{4!} - k^2(16 + 44k^2 + k^4)\frac{z^6}{6!} + \cdots$.

3. Prove $\int_0^\phi \frac{d\phi}{\sqrt{1 - l^2\sin^2\phi}} = \frac{1}{l}\int_0^\psi \frac{d\psi}{\sqrt{1 - l^{-2}\sin^2\psi}}$, $\quad l > 1, \quad \sin^2\psi = l^2\sin^2\phi$.

4. Carry out the computations in these cases:

(α) $\int_0^\phi \frac{d\theta}{\sqrt{1 - 0.1\sin^2\theta}}$ to find K, $\quad \mathrm{sn}\,\frac{2}{3}K, \quad F\left(\frac{1}{8}\pi, \frac{1}{\sqrt{10}}\right)$.

(β) $\int_0^\phi \frac{d\theta}{\sqrt{1 - 0.9\sin^2\theta}}$ to find K, $\quad \mathrm{sn}\,\frac{1}{3}K, \quad F\left(\frac{1}{3}\pi, \frac{3}{\sqrt{10}}\right)$.

5. A pendulum oscillates through an angle of (α) 180°, (β) 90°, (γ) 340°. Find the periodic time, the position at $t = \frac{2}{3}K$, and the time at which the pendulum makes an angle of 30° with the vertical.

6. With the aid of Ex. 3 find the arc of the lemniscate $r^2 = 2a^2 \cos 2\phi$. Also the arc from $\phi = 0$ to $\phi = 30°$, and the middle point of the arc.

7. A bead moves around a vertical circle. The velocity at the top is to the velocity at the bottom as $1:n$. Express the solution in terms of elliptic functions.

8. In Ex. 7 compute the periodic time if $n = 2$, 3, or 10.

9. Neglecting gravity, solve the problem of the jumping rope. Take the x-axis horizontal through the ends of the rope, and the y-axis vertical through one end. Remember that "centrifugal force" varies as the distance from the axis of rotation. The first and second integrations give

$$dx = \frac{a^2 dy}{\sqrt{(b^2 - y^2)^2 - a^2}}, \qquad y = \sqrt{b^2 - a^2}\,\mathrm{sn}\left(\frac{\sqrt{b^2 + a^2}\,x}{a^2},\ \sqrt{\frac{b^2 - a^2}{b^2 + a^2}}\right).$$

10. Express $\displaystyle\int \frac{d\theta}{\sqrt{a - \cos\theta}}$, $a > 1$, in terms of elliptic functions.

11. A ladder stands on a smooth floor and rests at an angle of 30° against a smooth wall. Discuss the descent of the ladder after its release from this position. Find the time which elapses before the ladder leaves the wall.

12. A rod is placed in a smooth hemispherical bowl and reaches from the bottom of the bowl to the edge. Find the time of oscillation when the rod is released.

190. Legendre's Integrals II and III. The treatment of

$$\int_0^w \frac{\sqrt{1 - k^2 w^2}}{\sqrt{1 - w^2}}\,dw = \int_0^w \frac{(1 - k^2 w^2)\,dw}{\sqrt{(1 - w^2)(1 - k^2 w^2)}} \tag{II}$$

by the method of conformal mapping to determine the function and its inverse does not give satisfactory results, for the map of the Riemann surface on the z-plane is not a simple region. But the integral may be treated by a change of variable and be reduced to the integral of an elliptic function. For with $w = \mathrm{sn}\,u$, $u = \mathrm{sn}^{-1} w$,

$$\begin{aligned}\int_0^w \frac{(1 - k^2 w^2)\,dw}{\sqrt{(1 - w^2)(1 - k^2 w^2)}} &= \int_0^u (1 - k^2 \mathrm{sn}^2 u)\,du \\ &= u - k^2 \int_0^u \mathrm{sn}^2 u\,du.\end{aligned} \tag{12}$$

The problem thus becomes that of integrating $\mathrm{sn}^2 u$. To effect the integration, $\mathrm{sn}^2 u$ will be expressed as a derivative.

The function $\mathrm{sn}^2 u$ is doubly periodic with periods $2K$, $2iK'$, and with a pole of the second order at $u = iK'$. But now

$$\Theta(u + 2K) = \Theta(u), \qquad \Theta(u + 2iK') = -q^{-1} e^{-\frac{i\pi}{K}u}\,\Theta(u)$$

$$\log\Theta(u + 2K) = \log\Theta(u), \quad \log(\Theta + 2iK') = \log\Theta(u) - \frac{i\pi}{K}u - \log(-q).$$

It then appears that the second derivative of $\log \Theta(u)$ also has the periods $2K$, $2iK'$. Introduce the zeta function

$$Z(u) = \frac{d}{du}\log\Theta(u) = \frac{\Theta'(u)}{\Theta(u)}, \qquad Z'(u) = \frac{d}{du}\frac{\Theta'(u)}{\Theta(u)}. \tag{13}$$

The expansion of $\Theta'(u)$ shows that $\Theta'(u) = 0$ at $u = mK$. About $u = iK'$ the expansions of $Z'(u)$ and $\operatorname{sn}^2 u$ are

$$Z'(u) = -\frac{1}{(u-iK')^2} + a_0 + \cdots, \qquad \operatorname{sn}^2 u = \frac{1}{k^2}\frac{1}{(u-iK')^2} + b_0 + \cdots.$$

Hence $\qquad k^2 \operatorname{sn}^2 u = -Z'(u) + Z'(0), \qquad Z'(0) = \Theta''(0)/\Theta(0),$

and
$$k^2\int_0^u \operatorname{sn}^2 u\, du = -Z(u) + uZ'(0),$$

$$\int_0^u (1 - k^2\operatorname{sn}^2 u)\, du = u(1 - Z'(0)) + Z(u). \tag{14}$$

The derivation of the expansions of $Z'(u)$ and $\operatorname{sn}^2 u$ about $u = iK'$ are easy.

$$\Theta(u) = C\prod\left(1 - q^{2n+1}e^{\pm\frac{i\pi}{K}u}\right), \qquad \log\Theta(u) = \sum\log\left(1 - q^{2n+1}e^{\pm\frac{i\pi}{K}u}\right) + \log C$$

$$\log\Theta(u) = \log\left(1 - qe^{-\frac{i\pi}{K}u}\right) + \text{function analytic near } u = iK'.$$

$$\frac{\Theta'(u)}{\Theta(u)} = \frac{i\pi q e^{-\frac{i\pi}{K}u}}{K\left(1 - qe^{-\frac{i\pi}{K}u}\right)} + \cdots = \frac{i\pi q}{K\left(e^{\frac{i\pi}{K}u} - q\right)} + \cdots,$$

$$f(u) = e^{\frac{i\pi}{K}u} = f(iK') + (u - iK')f'(iK') + \cdots = q + (u - iK')\frac{i\pi}{K}q + \cdots,$$

$$\frac{\Theta'(u)}{\Theta(u)} = \frac{+1}{u - iK'} + \cdots, \qquad \frac{d}{du}\frac{\Theta'(u)}{\Theta(u)} = \frac{-1}{(u-iK')^2} + \cdots.$$

$$\operatorname{sn}(u + iK') = \frac{1}{k}\frac{1}{\operatorname{sn} u}, \qquad \operatorname{sn}^2(u + iK') = \frac{1}{k^2}\frac{1}{\operatorname{sn}^2 u},$$

$$f(u) = \operatorname{sn} u = uf'(0) + \tfrac{1}{6}u^3 f'''(0) + \cdots = u + cu^3 + \cdots,$$

$$\operatorname{sn}^2(u + iK') = \frac{1}{k^2}\frac{1}{\operatorname{sn}^2 u} = \frac{1}{k^2}\left(\frac{1}{u} - cu + \cdots\right)^2 = \frac{1}{k^2}\left(\frac{1}{u^2} - 2c + \cdots\right),$$

$$\operatorname{sn}^2 u = \frac{1}{k^2}\left(\frac{1}{(u-iK')^2} - 2c + \cdots\right).$$

In a similar manner may be treated the integral

$$\int_0^w \frac{dw}{(w^2 - \alpha)\sqrt{(1-w^2)(1-k^2w^2)}} = \int_0^u \frac{du}{\operatorname{sn}^2 u - \alpha}. \tag{III}$$

Let a be so chosen that $\operatorname{sn}^2 a = \alpha$. The integral becomes

$$\int_0^u \frac{du}{\operatorname{sn}^2 u - \operatorname{sn}^2 a} = \frac{1}{2\operatorname{sn} a \operatorname{cn} a \operatorname{dn} a}\int \frac{2\operatorname{sn} a\operatorname{cn} a\operatorname{dn} a}{\operatorname{sn}^2 u - \operatorname{sn}^2 a}\,du. \tag{15}$$

The integrand is a function with periods $2K$, $2iK'$ and with simple poles at $u = \pm a$. To find the residues at these poles note

$$\lim_{u \doteq \pm a} \frac{u \mp a}{\operatorname{sn}^2 u - \operatorname{sn}^2 a} = \lim_{u \doteq \pm a} \frac{1}{2 \operatorname{sn} u \operatorname{cn} u \operatorname{dn} u} = \frac{\pm 1}{2 \operatorname{sn} a \operatorname{cn} a \operatorname{dn} a}.$$

The coefficient of $(u \mp a)^{-1}$ in expanding about $\pm a$ is therefore ± 1. Such a function may be written down. In fact

$$\begin{aligned}\frac{2 \operatorname{sn} a \operatorname{cn} a \operatorname{dn} a}{\operatorname{sn}^2 u - \operatorname{sn}^2 a} &= \frac{H'(u-a)}{H(u-a)} - \frac{H'(u+a)}{H(u+a)} + C \\ &= \mathrm{Z}_1(u-a) - \mathrm{Z}_1(u+a) + C,\end{aligned}$$

if $Z_1 = H'/H$. The verification is as above. To determine C let $u = 0$.

Then $\qquad C = -\dfrac{2 \operatorname{cn} a \operatorname{dn} a}{\operatorname{sn} a} + 2\,\mathrm{Z}_1(a),\quad$ but $\quad \operatorname{sn} u = \dfrac{1}{\sqrt{k}} \dfrac{H(u)}{\Theta(u)},$

and $\qquad \dfrac{d}{du} \log \operatorname{sn} u = \dfrac{\operatorname{cn} u \operatorname{dn} u}{\operatorname{sn} u} = \mathrm{Z}_1(u) - \mathrm{Z}(u).$

Hence C reduces to $2\,\mathrm{Z}(a)$ and the integral is

$$\int_0^u \frac{du}{\operatorname{sn}^2 u - \operatorname{sn}^2 a} = \frac{1}{2 \operatorname{sn} a \operatorname{cn} a \operatorname{dn} a}\left[\log \frac{H(a-u)}{H(a+u)} + 2\,u\mathrm{Z}(a)\right]. \quad (16)$$

The integrals here treated by the substitution $w = \operatorname{sn} u$ and thus reduced to the integrals of elliptic functions are but special cases of the integration of any rational function $R(w, \sqrt{W})$ of w and the radical of the biquadratic $W = (1 - w^2)(1 - k^2 w^2)$. The use of the substitution is analogous to the use of $w = \sin u$ in converting an integral of $R(w, \sqrt{1 - w^2})$ into an integral of trigonometric functions. Any rational function $R(w, \sqrt{W})$ may be written, by rationalization, as

$$\begin{aligned}R(w, \sqrt{W}) &= \frac{R(w) + R(w)\sqrt{W}}{R(w) + R(w)\sqrt{W}} = \frac{R(w) + R(w)\sqrt{W}}{R(w)} \\ &= R_1(w) + \frac{R(w)}{\sqrt{W}} = R_1(w) + \frac{w R_2(w^2) + R_3(w^2)}{\sqrt{W}}\end{aligned}$$

where R means not always the same function. The integral of $R(w, \sqrt{W})$ is thus reduced to the integral of $R_1(w)$ which is a rational fraction, plus the integral of $wR_2(w^2)/\sqrt{W}$ which by the substitution $w^2 = u$ reduces to an integral of $R(u, \sqrt{(1-u)(1-k^2u)})$ and may be considered as belonging to elementary calculus, plus finally

$$\int \frac{R_3(w^2)}{\sqrt{W}}\, dw = \int R_3(\operatorname{sn}^2 u)\, du, \qquad w = \operatorname{sn} u.$$

By the method of partial fractions R_3 may be resolved and

$$\int \operatorname{sn}^{2n} u\, du \qquad n \gtreqless 0, \qquad \int \frac{du}{(\operatorname{sn}^2 u - \alpha)^n} \qquad n > 0$$

are the types of integrals which must be evaluated to finish the integration of the given $R(w, \sqrt{W})$. An integration by parts (B. O. Peirce, No. 567) shows that for

the first type n may be lowered if positive and raised if negative until the integral is expressed in terms of the integrals of $\mathrm{sn}^2 x$ and $\mathrm{sn}^0 x = 1$, of which the first is integrated above. The second type for any value of n may be obtained from the integral for $n = 1$ given above by differentiating with respect to α under the sign of integration. Hence the whole problem of the integration of $R(w, \sqrt{W})$ may be regarded as solved.

191. With the substitution $w = \sin\phi$, the integral II becomes

$$E(\phi, k) = \int_0^\phi \sqrt{1 - k^2 \sin^2\theta}\, d\theta = \int_0^w \frac{\sqrt{1 - k^2 w^2}}{\sqrt{1 - w^2}}\, dw \tag{17}$$
$$= u(1 - \mathrm{Z}'(0)) + \mathrm{Z}(u), \qquad u = F(\phi, k).$$

In particular $E(\frac{1}{2}\pi, k)$ is called the complete integral of the second kind and is generally denoted by E. When $\phi = \frac{1}{2}\pi$, the integral $u = F(\phi, k)$ becomes the complete integral K. Then

$$E = K(1 - \mathrm{Z}'(0)) + \mathrm{Z}(K) = K(1 - \mathrm{Z}'(0)), \tag{18}$$

and
$$E(\phi, k) = EF(\phi, k)/K + \mathrm{Z}(u). \tag{19}$$

The problem of computing $E(\phi, k)$ thus reduces to that of computing K, E, $F(\phi, k) = u$, and $\mathrm{Z}(u)$. The methods of obtaining K and $F(\phi, k)$ have been given. The series for $\mathrm{Z}(u)$ converges rapidly. The value of E may be found by computing $K(1 - \mathrm{Z}'(0))$.

For the convenience of logarithmic computation note that

$$\frac{K - E}{K} = \mathrm{Z}'(0) = \frac{\Theta''(0)}{\Theta(0)} = \sqrt{\frac{\pi}{2Kk'}} \cdot \frac{2\pi^2}{K^2}(q - 4q^4 + 9q^9 - \cdots)$$

or
$$K - E = \tfrac{1}{2}\pi/\sqrt{k'} \cdot (2\pi/K)^{\frac{3}{2}} q(1 - 4q^3 + \cdots). \tag{20}$$

Also
$$\mathrm{Z}(u) = \frac{\Theta'(u)}{\Theta(u)} = \frac{2q\pi}{K} \frac{\sin 2v - 2q^3 \sin 4v + \cdots}{1 - 2q\cos 2v + 2q^4 \cos 4v - \cdots} \tag{21}$$

where $v = \pi u/2K$. These series neglect only terms in q^8, which will barely affect the fifth place when $k \leqq \sin 82°$ or $k^2 \leqq 0.98$. The series as written therefore cover most of the cases arising in practice. For instance in the problem which gives the name to the elliptic functions and integrals, the problem of finding the arc of the ellipse $x = a\sin\phi$, $y = b\cos\phi$,

$$ds = \sqrt{a^2\cos^2\phi + b^2\sin^2\phi}\, d\phi = a\sqrt{1 - e^2\sin^2\phi}\, d\phi;$$

the eccentricity e may be as high as 0.99 without invalidating the approximate formulas. An example follows.

Let it be required to determine the length of the quadrant of an ellipse of eccentricity $e = 0.9$ and also the length of the portion over half the semiaxis major. Here the series in q' converge better than those in q, but as the proper

expression to replace $Z(u)$ has not been found, it will be more convenient to use the series in q and take an additional term or two. As $k = 0.9$, $k'^2 = 0.19$.

$\operatorname{Lg} k'^2 = 9.27875$	$\operatorname{Lg}(1-\sqrt{k'}) = 9.53120$	$5 \text{ diff.} = 6.55515$
$\operatorname{Lg}\sqrt{k'} = 9.81969$	$\operatorname{Lg}(1+\sqrt{k'}) = 0.22017$	$\operatorname{Lg} 16 = 1.20412$
$\sqrt{k'} = 0.66022$	$\text{diff.} = 9.31103$	$\operatorname{Lg} \text{term } 2 = 5.35103$
$1-\sqrt{k'} = 0.33978$	$\operatorname{Lg} 2 = 0.30103$	$\text{term } 1 = 0.10233$
$1+\sqrt{k'} = 1.66022$	$\operatorname{Lg} \text{term } 1 = 9.01000$	$\text{term } 2 = 0.00002$
		$q = 0.10235.$

$\operatorname{Lg} q = 9.0101$	$\operatorname{Lg} 2\pi = 0.7982$	$\operatorname{Lg} \frac{1}{2}\pi/\sqrt{k'} = 0.3764$
$3 \operatorname{Lg} q = 7.0303$	$-2\operatorname{Lg}(1+\sqrt{k'}) = 9.5597$	$\frac{3}{2}\log 2\pi/K = 0.6603$
$4 \operatorname{Lg} q = 6.0404$	$\operatorname{Lg}(1+2q^4) = 0.0001$	$\operatorname{Lg} q = 9.0101$
$q^3 = 0.0011$	$\operatorname{Lg} K = 0.3580$	$\operatorname{Lg}(1-4q^3) = 9.9981$
$q^4 = 0.0001$	$K = 2.280$	$\operatorname{Lg}(K-E) = 0.0449.$

Hence $K - E = 1.109$ and $E = 1.171$. The quadrant is $1.171\,a$. The point corresponding to $x = \frac{1}{2}a$ is given by $\phi = 30°$. Then $\operatorname{dn} F = \sqrt{1 - 0.2025}$.

$\operatorname{Lg} \operatorname{dn} F = 9.9509$	$\frac{1}{4}\pi - \lambda = 8°31\frac{1}{2}'$	$\cos 2\nu = 0.7323$
$\operatorname{Lg}\sqrt{k'} = 9.8197$	$\operatorname{Lg} \tan = 9.1758$	Hence 4ν near $90°$
$\operatorname{Lg} \cot\lambda = 0.1312$	$\operatorname{Lg} 2q = 9.3111$	$1 + 2q^4\cos 4\nu = 1.0000$
$\lambda = 36°28\frac{1}{2}'$	$\operatorname{Lg}\cos 2\nu = 9.8647$	$2\nu = 42°55'.$

Now $180\,F = K(42.92)$. The computation for F, Z, $E(\frac{1}{6}\pi)$ is then

$\operatorname{Lg} K = 0.3580$	$\operatorname{Lg} 2\pi/K = 0.4402$	$\operatorname{Lg} E/K = 9.7106$
$\operatorname{Lg} 42.92 = 1.6326$	$\operatorname{Lg} q = 9.0101$	$\operatorname{Lg} F = 9.7353$
$-\operatorname{Lg} 180 = 7.7447$	$\operatorname{Lg}\sin 2\nu = 9.8331$	$EF/K = 0.2792$
$\operatorname{Lg} F = 9.7353$	$-\operatorname{Lg}(1 - 2q\cos 2\nu) = 0.0705$	$Z = 0.2256\,*$
$F = 0.5436$	$\operatorname{Lg} Z = 9.3539$	$E(\frac{1}{6}\pi) = 0.5048.$

The value of Z marked * is corrected for the term $-2q^3 \sin 4\nu$. The part of the quadrant over the first half of the axis is therefore $0.5048\,a$ and $0.666\,a$ over the second half. To insure complete four-figure accuracy in the result, five places should have been carried in the work, but the values here found check with the table except for one or two units in the last place.

EXERCISES

1. Prove the following relations for $Z(u)$ and $Z_1(u)$.

$$Z(-u) = -Z(u), \qquad Z(u + 2K) = Z(u), \qquad Z(u + 2iK') = Z(u) - i\pi/K.$$

If $$Z_1(u) = \frac{d}{du}\log H(u) = \frac{H'(u)}{H(u)}, \qquad Z_1(u + iK') = Z(u) - \frac{i\pi}{2K},$$

$$\frac{1}{\operatorname{sn}^2 u} = -Z_1'(u) + Z'(0), \qquad \int \frac{du}{\operatorname{sn}^2 u} = -Z_1(u) + uZ'(0),$$

$$Z_1(u) - Z(u) = \frac{d}{du}\log \operatorname{sn} u = \frac{\operatorname{cn} u \operatorname{dn} u}{\operatorname{sn} u}, \qquad Z_1(0) = \infty.$$

2. An elliptic function with periods $2K$, $2iK'$ and simple poles at $a_1, a_2, \cdots, a_n$ with residues $c_1, c_2, \cdots, c_n$, $\Sigma c = 0$, may be written

$$f(u) = c_1 \mathrm{Z}_1(u - a_1) + c_2 \mathrm{Z}_1(u - a_2) + \cdots + c_n \mathrm{Z}_1(u - a_n) + \text{const.}$$

3. $\dfrac{k^2 \operatorname{sn} a \operatorname{cn} a \operatorname{dn} a \operatorname{sn}^2 u}{1 - k^2 \operatorname{sn}^2 a \operatorname{sn}^2 u} = \dfrac{1}{2}\mathrm{Z}(u - a) - \dfrac{1}{2}\mathrm{Z}(u + a) + \mathrm{Z}'(a),$

$$k^2 \operatorname{sn} a \operatorname{cn} a \operatorname{dn} a \int_0^u \frac{\operatorname{sn}^2 u \operatorname{dn}}{1 - k^2 \operatorname{sn}^2 a \operatorname{sn}^2 u} = \frac{1}{2}\log\frac{\Theta(a - u)}{\Theta(a + u)} + u\mathrm{Z}'(a).$$

4. (α) $\displaystyle\int \frac{\lambda du}{\operatorname{sn}^2 \sqrt{\lambda} u} = \lambda u Z'(0) - \sqrt{\lambda}\, \mathrm{Z}(\sqrt{\lambda} u) - \sqrt{\lambda}\frac{\operatorname{cn}\sqrt{\lambda} u \operatorname{dn}\sqrt{\lambda} u}{\operatorname{sn}\sqrt{\lambda} u} + C$

$$- \lambda u - \sqrt{\lambda} E(\phi = \sin^{-1}\operatorname{sn}\sqrt{\lambda} u) - \sqrt{\lambda}\frac{\operatorname{cn}\sqrt{\lambda} u \operatorname{dn}\sqrt{\lambda} u}{\operatorname{sn}\sqrt{\lambda} u} + C,$$

(β) $\displaystyle\int \frac{k'^2 du}{\operatorname{dn}^2 u} = \int \operatorname{dn}^2 u du - k^2 \frac{\operatorname{sn} u \operatorname{cn} u}{\operatorname{dn} u} = E(\phi = \sin^{-1}\operatorname{sn} u) - k^2 \frac{\operatorname{sn} u \operatorname{cn} u}{\operatorname{dn} u},$

(γ) $\displaystyle\int \frac{\operatorname{cn}^2 u du}{\operatorname{sn}^2 u \operatorname{dn}^2 u} = u - 2E(\phi = \sin^{-1}\operatorname{sn} u) + \frac{\operatorname{cn} u}{\operatorname{sn} u \operatorname{dn} u}(1 - 2\operatorname{dn}^2 u).$

5. Find the length of the quadrant and of the portion of it cut off by the latus rectum in ellipses of eccentricity $e = 0.1, 0.5, 0.75, 0.95$.

6. If e is the eccentricity of the hyperbola $x^2/a^2 - y^2/b^2 = 1$, show that

$$s = \frac{b^2}{ae}\int_0^\phi \frac{\sec^2\phi d\phi}{\sqrt{1 - k^2\sin^2\phi}}, \qquad \text{where } \frac{ae}{b^2}y = \tan\phi, \qquad k = \frac{1}{e},$$

$$= \frac{b^2}{ae}F(\phi, k) - aeE(\phi, k) + ae\tan\phi\sqrt{1 - k^2\sin^2\phi}.$$

7. Find the arc of the hyperbola cut off by the latus rectum if $e = 1.2, 2, 3$.

8. Show that the length of the jumping rope (Ex. 9, p. 511) is

$$a(k'K/\sqrt{2} + \sqrt{2}E/k').$$

9. A flexible trough is filled with water. Find the expression of the shape of a cross section of the trough in terms of $F(\phi, k)$ and $E(\phi, k)$.

10. If an ellipsoid has the axes $a > b > c$, find the area of one octant.

$$\frac{1}{4}\pi c^2 + \frac{\pi ab}{4\sin\phi}\left[\frac{c^2}{a^2}F(\phi, k) + \frac{a^2 - c^2}{a^2}E(\phi, k)\right], \qquad \cos\phi = \frac{c}{a}, \qquad k^2 = \frac{b^2 - c^2}{b^2\sin^2\phi}.$$

11. Compute the area of the ellipsoid with axes 3, 2, 1.

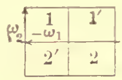
12. A hole of radius b is bored through a cylinder of radius $a > b$ centrally and perpendicularly to the axis. Find the volume cut out.

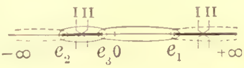
13. Find the area of a right elliptic cone, and compute the area if the altitude is 3 and the semiaxes of the base are $1\frac{1}{2}$ and 1.

192. Weierstrass's integral and its inversion. In studying the general theory of doubly periodic functions (§ 182), the two special functions $p(u)$, $p'(u)$ were constructed and discussed. It was seen that

$$\begin{aligned} z &= \int_\infty^w \frac{dw}{\sqrt{4w^3 - g_2 w - g_3}}, \qquad w = p(z), \qquad \infty = p(0), \\ &= \int_\infty^w \frac{dw}{\sqrt{4(w-e_1)(w-e_2)(w-e_3)}}, \qquad e_1 + e_2 + e_3 = 0, \end{aligned} \tag{22}$$

where the fixed limit ∞ has been added to the integral to make $w = \infty$ and $z = 0$ correspond and where the roots have been called e_1, e_2, e_3. Conversely this integral could be studied in detail by the method of mapping; but the method to be followed is to make only cursory use of the conformal map sufficient to give a hint as to how the function $p(z)$ may be expressed in terms of the functions sn z and cn z. The discussion will be restricted to the case which arises in practice, namely, when g_2 and g_3 are real quantities. There are two cases to consider, one when all three roots are real, the other when one is real and the other two are conjugate imaginary. The root e_1 will be taken as the largest real root, and e_2 as the smallest root if all three are real. Note that the sum of the three is zero.





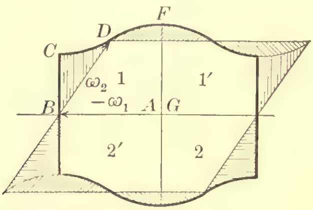
In the case of three real roots the Riemann surface may be drawn with junction lines e_2, e_3, and e_1, ∞. The details of the map may readily be filled in, but the observation is sufficient that there are only two essentially different paths closed on the surface, namely, about e_2, e_3 (which by deformation is equivalent to one about e_1, ∞) and about e_3, e_1 (which is equivalent to one about e_2, $-\infty$). The integral about e_2, e_3 is real and will be denoted by $2\,\omega_1$, that about e_3, e_1 is pure imaginary and will be denoted by $2\,\omega_2$. If the function $p(z)$ be constructed as in § 182 with $\omega = 2\,\omega_1$, $\omega' = 2\,\omega_2$ the function will have as always a double pole at $z = 0$. As the periods are real and pure imaginary, it is natural to try to express $p(z)$ in terms of sn z. As $p(z)$ depends on two constants g_2, g_3, whereas sn z depends on only the one k, the function $p(z)$ will be expressed in terms of $\text{sn}(\sqrt{\lambda} z, k)$, where the two constants λ, k are to be determined so as to fulfill the identity $p'^2 = 4\,p^3 - g_2 p - g_3$. In particular try

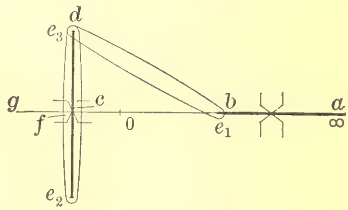
$$p(z) = A + \frac{\lambda}{\text{sn}^2(\sqrt{\lambda} z, k)}, \qquad A, \lambda, k \text{ constants.}$$

This form surely gives a double pole at $z = 0$ with the expansion $1/z^2$. The determination is relegated to the small text. The result is

$$p(z) = e_2 + \frac{e_1 - e_2}{\mathrm{sn}^2(\sqrt{\lambda}z, k)}, \qquad k^2 = \frac{e_3 - e_2}{e_1 - e_2} < 1,$$
$$\lambda = e_1 - e_2 > 0, \qquad \omega_1\sqrt{\lambda} = K, \qquad \omega_2\sqrt{\lambda} = iK'. \tag{23}$$

In the case of one real and two conjugate imaginary roots, the Riemann surface may be drawn in a similar manner. There are again two independent closed paths, one about e_2, e_3 and another about e_3, e_1. Let the integrals about these paths be respectively $2\,\omega_1$ and $2\,\omega_2$. That





$2\,\omega_1$ is real may be seen by deforming the path until it consists of a very distant portion along which the integral is infinitesimal and a path in and out along e_1, ∞, which gives a real value to the integral. As $2\,\omega_2$ is not known to be pure imaginary and may indeed be shown to be complex, it is natural to try to express $p(z)$ in terms of $\mathrm{cn}\, z$ of which one period is real and the other complex. Try

$$p(z) = A + \mu\,\frac{1 + \mathrm{cn}(2\sqrt{\mu}z, k)}{1 - \mathrm{cn}(2\sqrt{\mu}z, k)}.$$

This form surely gives a double pole at $z = 0$ with the expansion $1/z^2$. The determination is relegated to the small text. The result is

$$p(z) = e_1 + \mu\,\frac{1 + \mathrm{cn}(2\sqrt{\mu}z, k)}{1 - \mathrm{cn}(2\sqrt{\mu}z, k)}, \qquad k^2 = \frac{1}{2} - \frac{3\,e_1}{4\,\mu} < 1,$$
$$\mu^2 = (e_1 - e_2)(e_1 - e_3), \qquad \sqrt{\mu}\omega_1 = K, \qquad \sqrt{\mu}\omega_2 = \tfrac{1}{2}(K + iK'). \tag{23′}$$

To verify these determinations, substitute in $p'^2 = 4p^3 - g_2p - g_3$.

$$p(z) = A + \frac{\lambda}{\mathrm{sn}^2(\sqrt{\lambda}z, k)}, \qquad p'(z) = -\frac{2\lambda^{\frac{3}{2}}}{\mathrm{sn}^3(\sqrt{\lambda}z, k)}\,\mathrm{cn}(\sqrt{\lambda}z, k)\,\mathrm{dn}(\sqrt{\lambda}z, k),$$

$$4\lambda^3\frac{(1-\mathrm{sn}^2)(1-k^2\mathrm{sn}^2)}{\mathrm{sn}^6} = 4\left(A^3 + \frac{3A^2\lambda}{\mathrm{sn}^2} + \frac{3A\lambda^2}{\mathrm{sn}^4} + \frac{\lambda^3}{\mathrm{sn}^6}\right) - g_2A - \frac{g_2\lambda}{\mathrm{sn}^2} - g_3.$$

Equate coefficients of corresponding powers of sn^2. Hence the equations

$$4A^3 - g_2A - g_3 = 0, \qquad 4\lambda^2k^2 = 12A^2 - g_2\lambda, \qquad -\lambda(1 + k^2) = 3A.$$

The first shows that A is a root e. Let $A = e_2$. Note $-g_2 = e_1e_2 + e_1e_3 + e_2e_3$.

$$\lambda \cdot \lambda k^2 = 3e_2^2 + e_1e_2 + e_1e_3 + e_2e_3 = (e_1 - e_2)(e_3 - e_2).$$

$$\lambda + \lambda k^2 = -3e_2 = e_1 - e_2 + e_3 - e_2.$$

by virtue of the relation $e_1 + e_2 + e_3 = 0$. The solution is immediate as given.

To verify the second determination, the substitution is similar.

$$p(z) = A + \mu\frac{1 + \mathrm{cn}\, 2\sqrt{\mu}z}{1 - \mathrm{cn}\, 2\sqrt{\mu}z}, \qquad p'(z) = -\frac{4\mu^{\frac{3}{2}}\,\mathrm{sn}\,\mathrm{dn}}{(1 - \mathrm{cn})^2}.$$

$$[p'(z)]^2 = 16\mu^3\frac{(1 + \mathrm{cn})(k'^2 + k^2\mathrm{cn}^2)}{(1 - \mathrm{cn})^3} = 4\mu^3[t^3 + 2(1 - 2k^2)t^2 + t]$$

where $t = (1 + \mathrm{cn})/(1 - \mathrm{cn})$. The identity $p'^2 = 4p^3 - g_2p - g_3$ is therefore

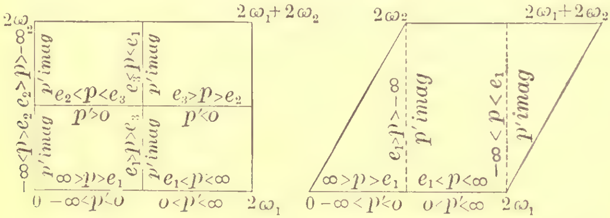
$$4\mu^3[t^3 + 2(1 - 2k^2)t^2 + t] = 4(A^3 + 3A^2\mu t + 3A\mu t^2 + \mu^3t^3) - g_2A - g_2\mu t - g_3.$$

$$4A^3 - g_2A - g_3 = 0, \qquad 4\mu^2 = 12A^2 - g_2, \qquad 2\mu(1 - 2k^2) = 3A.$$

Here let $A = e_1$. The solution then appears at once from the forms

$$\mu^2 = 3e_1^2 + e_1e_2 + e_1e_3 + e_2e_3 = (e_1 - e_2)(e_1 - e_3), \qquad \mu(1 - 2k^2) = 3A/2.$$

The expression of the function p in terms of the functions already studied permits the determination of the value of the function, and by inversion permits the solution of the equation $p(z) = c$. The function $p(z)$ may readily be expressed directly in terms of the theta series. In fact the periodic properties of the function and the corresponding properties of the quotients of theta series allow such a representation



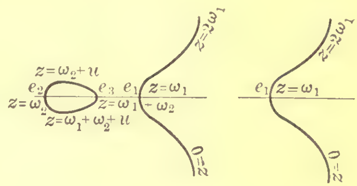
to be made from the work of § 175, provided the series be allowed complex values for q. But for practical purposes it is desirable to have the expression in terms of real quantities only, and this is the reason for a different expression in the two different cases here treated.*

The values of z for which $p(z)$ is real may be read off from (23) and (23′) or from the correspondence between the w-surface and the z-plane. They are indicated on the figures. The functions p and p' may be used to express parametrically the curve

$$y^2 = 4x^3 - g_2x - g_3 \quad \text{by} \quad y = p'(z), \qquad x = p(z).$$

* It is, however, *possible*, if desired, to transform the given cubic $4x^3 - g_2x - g_3$ with two complex roots into a similar cubic with all three roots real and thus avoid the duplicate forms. The transformation is not given here.

The figures indicate in the two cases the shape of the curves and the range of values of the parameter. As the function p is of the second order, the equation $p(z) = c$ has just two roots in the parallelogram, and as $p(z)$ is an even function, they will be of the form $z = a$ and $z = 2\omega_1 + 2\omega_2 - a$ and be symmetrically situated with respect to the center of the figure except in case a lies on the sides of the parallelogram so that $2\omega_1 + 2\omega_2 - a$ would lie on one of the excluded sides. The value of the odd function p' at these two points is equal and opposite. This corresponds precisely to the fact that to one value $x = c$ of x there are two equal and opposite values of y on the curve $y^2 = 4x^3 - g_2 x - g_3$. Conversely to each point of the parallelogram corresponds one point of the curve and to points symmetrically situated with respect to the center correspond points of the curve symmetrically situated with respect to the x-axis. Unless z is such as to make both $p(z)$ and $p'(z)$ real, the point on the curve will be imaginary.



193. The curve $y^2 = 4x^3 - g_2 x - g_3$ may be studied by means of the properties of doubly periodic functions. For instance

$$Ax + By + C = Ap'(z) + Bp(z) + C = 0$$

is the condition that the parameter z should be such that its representative point shall lie on the line $Ax + By + C = 0$. But the function $Ap'(z) + Bp(z) + C$ is doubly periodic with a pole of the third order; the function is therefore of the third order and there are just three points z_1, z_2, z_3 in the parallelogram for which the function vanishes. These values of z correspond to the three intersections of the line with the cubic curve. Now the roots of the doubly periodic function satisfy the relation

$$z_1 + z_2 + z_3 - 3 \times 0 = 2m_1\omega_1 + 2m_2\omega_2.$$

It may be observed that neither m_1 nor m_2 can be as great as 3. If conversely z_1, z_2, z_3 are three values of z which satisfy the relation $z_1 + z_2 + z_3 = 2m_1\omega_1 + 2m_2\omega_2$, the three corresponding points of the cubic will lie on a line. For if z_3' be the point in which a line through z_1, z_2 cuts the curve,

$$z_1 + z_2 + z_3' = 2m_1'\omega_1 + 2m_2'\omega_2, \qquad z_3 - z_3' = 2(m_1 - m_1')\omega_1 + 2(m_2 - m_2')\omega_2,$$

and hence z_3, z_3' are identical except for the addition of periods and must therefore be the same point on the parallelogram.

One application of this condition is to find the tangents to the curve from any point of the curve. Let z be the point from which and z' that to which the tangent is drawn. The condition then is $z + 2z' = 2m_1\omega_1 + 2m_2\omega_2$, and hence

$$z' = -\tfrac{1}{2}z, \qquad z' = -\tfrac{1}{2}z + \omega_1, \qquad z' = -\tfrac{1}{2}z + \omega_2, \qquad z' = -\tfrac{1}{2}z + \omega_1 + \omega_2$$

are the four different possibilities for z' corresponding to $m_1 = m_2 = 0$; $m_1 = 1$, $m_2 = 0$; $m_1 = 0$, $m_2 = 1$; $m_1 = 1$, $m_2 = 1$. To give other values to m_1 or m_2 would

merely reproduce one of the four points except for the addition of complete periods. Hence there are four tangents to the curve from any point of the curve. The question of the reality of these tangents may readily be treated. Suppose z denotes a real point of the curve. If the point lies on the infinite portion, $0 < z < 2\,\omega_1$, and the first two points z' will also satisfy the conditions $0 < z' < 2\,\omega_1$ except for the possible addition of $2\,\omega_1$. Hence there are always two real tangents to the curve from any point of the infinite branch. In case the roots e_1, e_2, e_3 are all real, the last two points z' will correspond to real points of the oval portion and all four tangents are real; in the case of two imaginary roots these values of z' give imaginary points of the curve and there are only two real tangents. If the three roots are real and z corresponds to a point of the oval, z is of the form $\omega_2 + u$ and all four values of z' are complex,

$$-\tfrac{1}{2}\,\omega_2 - \tfrac{1}{2}\,u, \qquad -\tfrac{1}{2}\,\omega_2 - \tfrac{1}{2}\,u + \omega_1, \qquad +\tfrac{1}{2}\,\omega_2 - \tfrac{1}{2}\,u, \qquad +\tfrac{1}{2}\,\omega_2 - \tfrac{1}{2}\,u + \omega_1,$$

and none of the tangents can be real. The discussion is complete.

As an inflection point is a point at which a line may cut a curve in three coincident points, the condition $3\,z = 2\,m_1\omega_1 + 2\,m_2\omega_2$ holds for the parameter z of such points. The possible different combinations for z are nine:

$$\begin{array}{lll} z = 0 & \tfrac{2}{3}\,\omega_2 & \tfrac{4}{3}\,\omega_2 \\ \tfrac{2}{3}\,\omega_1 & \tfrac{2}{3}\,\omega_1 + \tfrac{2}{3}\,\omega_2 & \tfrac{2}{3}\,\omega_1 + \tfrac{4}{3}\,\omega_2 \\ \tfrac{4}{3}\,\omega_1 & \tfrac{4}{3}\,\omega_1 + \tfrac{2}{3}\,\omega_2 & \tfrac{4}{3}\,\omega_1 + \tfrac{4}{3}\,\omega_2. \end{array}$$

Of these nine inflections only the three in the first column are real. When any two inflections are given a third can be found so that $z_1 + z_2 + z_3$ is a complete period, and hence the inflections lie three by three on twelve lines.

If p and p' be substituted in $Ax^2 + Bxy + Cy^2 + Dx + Ey + F$, the result is a doubly periodic function of order 6 with a pole of the 6th order at the origin. The function then has 6 zeros in the parallelogram connected by the relation

$$z_1 + z_2 + z_3 + z_4 + z_5 + z_6 = 2\,m_1\omega_1 + 2\,m_2\omega_2,$$

and this is the condition which connects the parameters of the 6 points in which the cubic is cut by the conic $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$. One application of interest is to the discussion of the conics which may be tangent to the cubic at three points z_1, z_2, z_3. The condition then reduces to $z_1 + z_2 + z_3 = m_1\omega_1 + m_2\omega_2$. If m_1, m_2 are 0 or any even numbers, this condition expresses the fact that the three points lie on a line and is therefore of little interest. The other possibilities, apart from the addition of complete periods, are

$$z_1 + z_2 + z_3 = \omega_1, \qquad z_1 + z_2 + z_3 = \omega_2, \qquad z_1 + z_2 + z_3 = \omega_1 + \omega_2.$$

In any of the three cases two points may be chosen at random on the cubic and the third point is then fixed. Hence there are three conics which are tangent to the cubic at any two assigned points and at some other point. Another application of interest is to the conics which have contact of the 5th order with the cubic. The condition is then $6\,z = 2\,m_1\omega_1 + 2\,m_2\omega_2$. As m_1, m_2 may have any of the 6 values from 0 to 5, there are 36 points on the cubic at which a conic may have contact of the 5th order. Among these points, however, are the nine inflections obtained by giving m_1, m_2 even values, and these are of little interest because the conic reduces to the inflectional tangent taken twice. There remain 27 points at which a conic may have contact of the 5th order with the cubic.

EXERCISES

1. The function $\zeta(z)$ is defined by the equation

$$-\zeta'(z) = p(z) \quad \text{or} \quad \zeta(z) = -\int p(z)\,dz = \frac{1}{z} - \frac{1}{3}c_1 z^3 + \cdots.$$

Show by Ex. 4, p. 516, that the value of ζ in the two cases is

$$\zeta(z) = -e_1 z + \sqrt{\lambda}E(\phi, k) + \sqrt{\lambda}\frac{\operatorname{cn}\sqrt{\lambda}z \operatorname{dn}\sqrt{\lambda}z}{\operatorname{sn}\sqrt{\lambda}z},$$

$$\zeta(z) = -(\mu + e_1)z + 2\sqrt{\mu}E(\phi, k) + \sqrt{\mu}\frac{\operatorname{cn}\sqrt{\mu}z}{\operatorname{sn}\sqrt{\mu}z \operatorname{dn}\sqrt{\mu}z}\left(2\operatorname{dn}^2\sqrt{\mu}z - 1\right),$$

where $\quad \lambda = e_1 - e_2, \quad k^2 = (e_3 - e_2)/(e_1 - e_2), \quad \phi = \sin^{-1}\operatorname{sn}\sqrt{\lambda}z,$

and $\quad \mu = \sqrt{(e_1 - e_2)(e_1 - e_3)}, \quad k^2 = \frac{1}{2} - 3e_1/4\mu, \quad \phi = \sin^{-1}\operatorname{sn}\sqrt{\mu}z.$

2. In case the three roots are real show that $p(z) - e_i$ is a square.

$$\sqrt{p(z) - e_1} = \sqrt{\lambda}\frac{\operatorname{cn}\sqrt{\lambda}z}{\operatorname{sn}\sqrt{\lambda}z}, \quad \sqrt{p(z) - e_2} = \frac{\sqrt{\lambda}}{\operatorname{sn}\sqrt{\lambda}z}, \quad \sqrt{p(z) - e_3} = \sqrt{\lambda}\frac{\operatorname{dn}\sqrt{\lambda}z}{\operatorname{sn}\sqrt{\lambda}z}$$

What happens in case there is only one real root?

3. Let $p(z\,;\,g_2, g_3)$ denote the function p corresponding to the radical

$$\sqrt{4p^3 - g_2 p - g_3}.$$

Compute $p(\frac{1}{2}\,;\,1, 0)$, $p(\frac{1}{4}\,;\,0, \frac{1}{2})$, $p(\frac{3}{4}\,;\,13, 6)$. Solve $p(z\,;\,1, 0) = 2$, $p(z\,;\,0, \frac{1}{2}) = 3$, $p(z\,;\,13, 6) = 10$.

4. If 6 of the 9 points in which a cubic cuts $y^2 = 4x^3 - g_2 x - g_3$ are on a conic, the other three are in a straight line.

5. If a conic has contact of the second order with the cubic at two points, the points of contact lie on a line through one of the inflections.

6. How many of the points at which a conic may have contact of the 5th order with the cubic are real? Locate the points at least roughly.

7. If a conic cuts the cubic in four fixed and two variable points, the line joining the latter two passes through a fixed point of the cubic.

8. Consider the space curve $x = \operatorname{sn} t$, $y = \operatorname{cn} t$, $z = \operatorname{dn} t$. Show that to each point of the rectangle $4K$ by $4iK'$ corresponds one point of the curve and conversely. Show that the curve is the intersection of the cylinders $x^2 + y^2 = 1$ and $k^2x^2 + z^2 = 1$. Show that a plane cuts the curve in 4 points and determine the relation between the parameters of the points.

9. How many osculating planes may be drawn to the curve of Ex. 8 from any point on it? At how many points may a plane have contact of the 3d order with the curve and where are the points?

10. In case the roots are real show that $\zeta(z)$ has the form

$$\zeta(z) = \frac{\eta_1}{\omega_1}z + \sqrt{\lambda}Z_1(\sqrt{\lambda}z), \qquad \eta_1 = \sqrt{\lambda}E - \frac{Ke_1}{\sqrt{\lambda}}.$$

Hence $$\log \sigma(z) = \int \zeta(z)\,dz = \frac{1}{2}\frac{\eta_1}{\omega_1} z^2 + \log H(\sqrt{\lambda} z) + C$$

or $$\sigma(z) = Ce^{\frac{1}{2}\frac{\eta_1}{\omega_1} z^2} H(\sqrt{\lambda} z).$$

11. By general methods like those of § 190 prove that

$$\frac{1}{p(z) - p(a)} = -\frac{1}{p'(a)}\left[\zeta(z+a) - \zeta(z-a) - 2\,\zeta(a)\right],$$

and $$\int \frac{dz}{p(z) - p(a)} = -\frac{1}{p'(a)} \log \frac{\sigma(z+a)}{\sigma(z-a)} + 2\frac{z\zeta(a)}{p'(a)}.$$

12. Let the functions θ be defined by these relations:

$$\theta(z) = H\left(\frac{Ku}{\omega_1}\right), \quad \theta_1(z) = H_1\left(\frac{Ku}{\omega_1}\right), \quad \theta_2(z) = \Theta\left(\frac{Ku}{\omega_1}\right), \quad \theta_3(z) = \Theta_1\left(\frac{Ku}{\omega_1}\right)$$

with $q = e^{\frac{\pi i \omega_2}{\omega_1}}$. Show that the θ-series converge if ω_1 is real and ω_2 is pure imaginary or complex with its imaginary part positive. Show more generally that the series converge if the angle from ω_1 to ω_2 is positive and less than 180°.

13. Let $$\sigma(z) = e^{\frac{1}{2}\frac{\eta_1}{\omega_1}z^2}\frac{\theta(z)}{\theta'(0)}, \qquad \sigma_a(z) = e^{\frac{1}{2}\frac{\eta_1}{\omega_1}z^2}\frac{\theta_a(z)}{\theta_a(0)}.$$

Prove $\sigma(z + 2\,\omega_1) = -e^{2\eta_1(z+\omega_1)}\sigma(z)$ and similar relations for $\sigma_a(z)$.

14. Let $$2\,\eta_2 = \frac{2\,\eta_1\omega_2}{\omega_1} - \frac{\pi i}{\omega_1}, \quad \text{or} \quad \eta_1\omega_2 - \eta_2\omega_1 = \frac{\pi i}{2}.$$

Prove $\sigma(z + 2\,\omega_2) = -e^{2\eta_2(z+\omega_2)}\sigma(z)$ and similar relations for $\sigma_a(z)$.

15. Show that $\sigma(-z) = -\sigma(z)$ and develop $\sigma(z)$ as

$$\sigma(z) = z + \left[\frac{\eta_1}{2\,\omega_1} + \frac{1}{6}\frac{\theta'''(0)}{\theta'(0)}\right]z^3 + \cdots = z + 0\cdot z^3 + \cdots, \quad \text{if} \quad \eta_1 = -\frac{\omega_1}{3}\frac{\theta'''(0)}{\theta'(0)}.$$

16. With the determination of η_1 as in Ex. 15 prove that

$$\frac{d}{dz}\log\sigma(z) = \zeta(z), \qquad -\frac{d^2}{dz^2}\log\sigma(z) = -\zeta'(z) = p(z)$$

by showing that $p(z)$ as here defined is doubly periodic with periods $2\,\omega_1$, $2\,\omega_2$, with a pole $1/z^2$ of the second order at $z = 0$ and with no constant term in its development. State why this identifies $p(z)$ with the function of the text.

CHAPTER XX

FUNCTIONS OF REAL VARIABLES

194. Partial differential equations of physics. In the solution of physical problems partial differential equations of higher order, particularly the second, frequently arise. With very few exceptions these equations are linear, and if they are solved at all, are solved by assuming the solution as a product of functions each of which contains only one of the variables. The determination of such a solution offers only a particular solution of the problem, but the combination of different particular solutions often suffices to give a suitably general solution. For instance

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \quad \text{or} \quad \frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \phi^2} = 0 \tag{1}$$

is Laplace's equation in rectangular and polar coördinates. For a solution in rectangular coördinates the assumption $V = X(x)\,Y(y)$ would be made, and the assumption $V = R(r)\,\Phi(\phi)$ for a solution in polar coördinates. The equations would then become

$$\frac{X''}{X} + \frac{Y''}{Y} = 0 \quad \text{or} \quad \frac{r^2R''}{R} + r\frac{R'}{R} + \frac{\Phi''}{\Phi} = 0. \tag{2}$$

Now each equation as written is a sum of functions of a single variable. But a function of x cannot equal a function of y and a function of r cannot equal a function of ϕ unless the functions are constant and have the same value. Hence

$$\frac{X''}{X} = -m^2, \qquad \frac{\Phi''}{\Phi} = -m^2,$$
$$\text{or} \tag{2'}$$
$$\frac{Y''}{Y} = +m^2, \qquad \frac{r^2R''}{R} + r\frac{R'}{R} = +m^2.$$

These are ordinary equations of the second order and may be solved as such. The second case will be treated in detail.

The solution corresponding to any value of m is

$$\Phi = a_m \cos m\phi + b_m \sin m\phi, \qquad R = A_m r^m + B_m r^{-m}$$

and $$V = R\Phi = (A_m r^m + B_m r^{-m})(a_m \cos m\phi + b_m \sin m\phi)$$

or $$V = \sum_m (A_m r^m + B_m r^{-m})(a_m \cos m\phi + b_m \sin m\phi). \tag{3}$$

That any number of solutions corresponding to different values of m may be added together to give another solution is due to the *linearity* of the given equation (§ 96). It may be that a single term will suffice as a solution of a given problem. But it may be seen in general that: A solution for V may be found in the form of a Fourier series which shall give V any assigned values on a unit circle and either be convergent for all values within the circle or be convergent for all values outside the circle. In fact let $f(\phi)$ be the values of V on the unit circle. Expand $f(\phi)$ into its Fourier series

$$f(\phi) = \tfrac{1}{2} a_0 + \sum_m (a_m \cos m\phi + b_m \sin m\phi).$$

Then $$V = \tfrac{1}{2} a_0 + \sum_m r^m (a_m \cos m\phi + b_m \sin m\phi) \tag{3'}$$

will be a solution of the equation which reduces to $f(\phi)$ on the circle and, as it is a power series in r, converges at every point within the circle. In like manner a solution convergent outside the circle is

$$V = \tfrac{1}{2} a_0 + \sum_m r^{-m} (a_m \cos m\phi + b_m \sin m\phi). \tag{3''}$$

The infinite series for V have been called solutions of Laplace's equation. As a matter of fact they have not been proved to be solutions. The finite sum obtained by taking any number of terms of the series would surely be a solution; but the limit of that sum when the series becomes infinite is not thereby proved to be a solution even if the series is convergent. For theoretical purposes it would be necessary to give the proof, but the matter will be passed over here as having a negligible bearing on the practical solution of many problems. For in practice the values of $f(\phi)$ on the circle could not be exactly known and could therefore be adequately represented by a finite and in general not very large number of terms of the development of $f(\phi)$, and these terms would give only a finite series for the desired function V.

In some problems it is better to keep the particular solutions separate, discuss each possible particular solution, and then imagine them compounded physically. Thus in the motion of a drumhead, the most general solution obtainable is not so instructive as the particular solution corresponding to particular notes; and in the motion of the surface of the ocean it is preferable to discuss individual types of waves and compound them according to the law of superposition of small vibrations (p. 226). For example if

$$\frac{1}{c^2}\frac{\partial^2 z}{\partial t^2} = \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}, \qquad \frac{1}{c^2}\frac{T''}{T} = \frac{X''}{X} + \frac{Y''}{Y}, \qquad z = XYT,$$

be taken as the equation of motion of a rectangular drumhead,

$$X = \begin{cases}\sin \alpha x, \\ \cos \alpha x,\end{cases} \qquad Y = \begin{cases}\sin \beta y, \\ \cos \beta y,\end{cases} \qquad T = \begin{cases}\sin c\sqrt{\alpha^2 + \beta^2}t \\ \cos c\sqrt{\alpha^2 + \beta^2}t\end{cases}$$

are particular solutions which may be combined in any way desired As the edges of the drumhead are supposed to be fixed at all times,

$$z = 0 \quad \text{if} \quad x = 0, \quad x = a, \quad y = 0, \quad y = b, \quad t = \text{anything},$$

where the dimensions of the head are a by b. Then the solution

$$z = XYT = \sin\frac{m\pi x}{a}\sin\frac{n\pi y}{b}\cos c\pi\sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}}\,t \tag{4}$$

is a possible type of vibration satisfying the given conditions at the perimeter of the head for any integral values of m, n. The solution is periodic in t and represents a particular note which may be omitted. A sum of such expressions multiplied by any constants would also be a solution and would represent a possible mode of motion, but would not be periodic in t and would represent no note.

195. For three dimensions Laplace's equation becomes

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right) = 0 \tag{5}$$

in polar coördinates. Substitute $V = R(r)\Theta(\theta)\Phi(\phi)$; then

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{1}{\Theta\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \frac{1}{\Phi\sin^2\theta}\frac{d^2\Phi}{d\phi^2} = 0.$$

Here the first term involves r alone and no other term involves r Hence the first term must be a constant, say, $n(n+1)$. Then

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - n(n+1)R = 0, \qquad R = Ar^n + Br^{-n-1}.$$

Next consider the last term after multiplying through by $\sin^2\theta$. It appears that $\Phi^{-1}\Phi''$ is a constant, say, $-m^2$. Hence

$$\Phi'' = -m^2\Phi, \qquad \Phi = a_m\cos m\phi + b_m\sin m\phi.$$

Moreover the equation for Θ now reduces to the simple form

$$\frac{d}{d\cos\theta}\left[(1-\cos^2\theta)\frac{d\Theta}{d\cos\theta}\right] + \left[n(n+1) - \frac{m^2}{1-\cos^2\theta}\right]\Theta = 0.$$

The problem is now separated into that of the integration of three differential equations of which the first two are readily integrable. The third equation is a generalization of Legendre's (Exs. 13–17, p. 252).

and in case n, m are positive integers the solution may be expressed in terms of polynomials $P_{n,m}(\cos\theta)$ in $\cos\theta$. Any expression

$$\sum_{n,m}(A_n r^n + B_n r^{-n-1})(a_m \cos m\phi + b_m \sin m\phi)P_{n,m}(\cos\theta)$$

is therefore a solution of Laplace's equation, and it may be shown that by combining such solutions into infinite series, a solution may be obtained which takes on any desired values on the unit sphere and converges for all points within or outside.

Of particular simplicity and importance is the case in which V is supposed independent of ϕ so that $m = 0$ and the equation for Θ is soluble in terms of Legendre's polynomials $P_n(\cos\theta)$ if n is integral. As the potential V of any distribution of matter attracting according to the inverse square of the distance satisfies Laplace's equation at all points exterior to the mass (§ 201), the potential of any mass symmetric with respect to revolution about the polar axis $\theta = 0$ may be expressed if its expression for points on the axis is known. For instance, the potential of a mass M distributed along a circular wire of radius a is

$$V = \frac{M}{\sqrt{a^2 + r^2}} = \begin{cases} \dfrac{M}{a}\left(1 - \dfrac{1}{2}\dfrac{r^2}{a^2} + \dfrac{1\cdot 3}{2\cdot 4}\dfrac{r^4}{a^4} - \dfrac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\dfrac{r^6}{a^6} + \cdots, \right. & r < a, \\ \dfrac{M}{a}\left(\dfrac{a}{r} - \dfrac{1}{2}\dfrac{a^3}{r^3} + \dfrac{1\cdot 3}{2\cdot 4}\dfrac{a^5}{r^5} - \dfrac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\dfrac{a^7}{r^7} + \cdots, \right. & r > a, \end{cases}$$

at a point distant r from the center of the wire along a perpendicular to the plane of the wire. The two series

$$V = \begin{cases} \dfrac{M}{a}\left(P_0 - \dfrac{1}{2}\dfrac{r^2}{a^2}P_2 + \dfrac{1\cdot 3}{2\cdot 4}\dfrac{r^4}{a^4}P_4 - \dfrac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\dfrac{r^6}{a^6}P_6 + \cdots, \right. & r < a, \\ \dfrac{M}{a}\left(\dfrac{a}{r}P_0 - \dfrac{1}{2}\dfrac{a^3}{r^3}P_2 + \dfrac{1\cdot 3}{2\cdot 4}\dfrac{a^5}{r^5}P_4 - \dfrac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\dfrac{a^7}{r^7}P_6 + \cdots, \right. & r > a, \end{cases}$$

are then precisely of the form $\Sigma A_n r^n P_n$, $\Sigma A_n r^{-n-1} P_n$ admissible for solutions of Laplace's equation and reduce to the known value of V along the axis $\theta = 0$ since $P_n(1) = 1$. They give the values of V at all points of space.

To this point the method of combining solutions of the given differential equations was to add them into a finite or infinite series. It is also possible to combine them by integration and to obtain a solution as a *definite integral* instead of as an infinite series. It should be noted in this case, too, that a limit of a sum has replaced a sum and that it would theoretically be necessary to demonstrate that the limit of the sum was really a solution of the given equation. It will be sufficient at this point to illustrate the method without any rigorous attempt to

justify it. Consider (2′) in rectangular coördinates. The solutions for X, Y are

$$\frac{X''}{X} = -m^2,\quad \frac{Y''}{Y} = m^2,\quad X = a_m \cos mx + b_m \sin mx,\quad Y = A_m e^{my} + B_m e^{-my},$$

where Y may be expressed in terms of hyperbolic functions. Now

$$\begin{aligned} V &= \int_{m_0}^{m_1} e^{-my}[a(m)\cos mx + b(m)\sin mx]\,dm \\ &= \lim \sum e^{-m_i y}[a(m_i)\cos m_i x + b(m_i)\sin m_i x]\,\Delta m_i \end{aligned} \tag{6}$$

is the limit of a sum of terms each of which is a solution of the given equation; for $a(m_i)$ and $b(m_i)$ are constants for any given value $m = m_i$, no matter what functions $a(m)$ and $b(m)$ are of m. It may be assumed that V is a solution of the given equation. Another solution could be found by replacing e^{-my} by e^{my}.

It is sometimes possible to determine $a(m)$, $b(m)$ so that V shall reduce to assigned values on certain lines. In fact (p. 466)

$$f(x) = \frac{1}{\pi}\int_0^{\infty}\int_{-\infty}^{+\infty} f(\lambda)\cos m(\lambda - x)\,d\lambda dm. \tag{7}$$

Hence if the limits for m be 0 and ∞ and if the choice

$$a(m) = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(\lambda)\cos m\lambda d\lambda, \qquad b(m) = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(\lambda)\sin m\lambda d\lambda$$

is taken for $a(m)$, $b(m)$, the expression (6) for V becomes

$$V = \frac{1}{\pi}\int_0^{\infty}\int_{-\infty}^{+\infty} e^{-my} f(\lambda)\cos m(\lambda - x)\,d\lambda dm \tag{8}$$

and reduces to $f(x)$ when $y = 0$. Hence a solution V is found which takes on any assigned values $f(x)$ along the x-axis. This solution clearly becomes zero when y becomes infinite. When $f(x)$ is given it is sometimes possible to perform one or more of the integrations and thus simplify the expression for V.

For instance if

$$f(x) = 1 \text{ when } x > 0 \quad\text{and}\quad f(x) = 0 \text{ when } x < 0,$$

the integral from $-\infty$ to 0 drops out and

$$V = \frac{1}{\pi}\int_0^{\infty}\int_0^{\infty} e^{-my}\cdot 1\cdot \cos m(\lambda - x)\,d\lambda dm = \frac{1}{\pi}\int_0^{\infty}\int_0^{\infty} e^{-my}\cos m(\lambda - x)\,dm d\lambda$$

$$= \frac{1}{\pi}\int_0^{\infty}\frac{y d\lambda}{y^2 + (\lambda - x)^2} = \frac{1}{\pi}\left(\frac{\pi}{2} + \tan^{-1}\frac{x}{y}\right) = 1 - \frac{1}{\pi}\tan^{-1}\frac{y}{x}.$$

It may readily be shown that when $y > 0$ the reversal of the order of integration is permissible; but as V is determined completely, it is simpler to substitute the value as found in the equation and see that $V''_{xx} + V''_{yy} = 0$, and to check the fact that V reduces to $f(x)$ when $y = 0$. It may perhaps be superfluous to state that the proved correctness of an answer does not show the justification of the steps by which that answer is found; but on the other hand as those steps were taken solely to obtain the answer, there is no practical need of justifying them if the answer is clearly right.

EXERCISES

1. Find the indicated particular solutions of these equations:

(α) $c^2\dfrac{\partial V}{\partial t} = \dfrac{\partial^2 V}{\partial x^2}$, $\quad V = \sum A_m e^{-m^2 t}(a_m \cos cmx + b_m \sin cmx)$,

(β) $\dfrac{1}{c^2}\dfrac{\partial^2 V}{\partial t^2} = \dfrac{\partial^2 V}{\partial x^2}$, $\quad V = \sum (A_m \cos cmt + B_m \sin cmt)(a_m \cos mx + b_m \sin mx)$,

(γ) $c^2\dfrac{\partial V}{\partial t} = \dfrac{\partial^2 V}{\partial x^2} + \dfrac{\partial^2 V}{\partial y^2}$, $\quad X = \begin{cases}\sin c\alpha x\\ \cos c\alpha x,\end{cases} \quad Y = \begin{cases}\sin c\beta y\\ \cos c\beta y,\end{cases} \quad T = e^{-(\alpha^2+\beta^2)t}$.

2. Determine the solutions of Laplace's equation in the plane that have $V = 1$ for $0 < \phi < \pi$ and $V = -1$ for $\pi < \phi < 2\pi$ on a unit circle.

3. If $V = |\pi - \phi|$ on the unit circle, find the expansion for V.

4. Show that $V = \Sigma a_m \sin m\pi x/l \cdot \cos cm\pi t/l$ is the solution of Ex. 1 (β) which vanishes at $x = 0$ and $x = l$. Determine the coefficients a_m so that for $t = 0$ the value of V shall be an assigned function $f(x)$. This is the problem of the violin string started from any assigned configuration.

5. If the string of Ex. 4 is started with any assigned velocity $\partial V/\partial t = f(x)$ when $t = 0$, show that the solution is $\Sigma a_m \sin m\pi x/l \cdot \sin cm\pi t/l$ and make the proper determination of the constants a_m.

6. If the drumhead is started with the shape $z = f(x, y)$, show that

$$z = \sum_{m,n} A_{m,n} \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b} \cos c\pi t\sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}},$$

$$A_{m,n} = \frac{4}{ab}\int_0^a\int_0^b f(x, y) \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b}\, dydx.$$

7. In hydrodynamics it is shown that $\dfrac{\partial^2 y}{\partial t^2} = \dfrac{g}{b}\dfrac{\partial}{\partial x}\left(hb\dfrac{\partial y}{\partial x}\right)$ is the differential equation for the surface of the sea in an estuary or on a beach of breadth b and depth h measured perpendicularly to the x-axis which is supposed to run seaward. Find

(α) $y = AJ_0(kx)\cos nt$, $\quad k^2 = n^2/gh$, $\qquad (\beta)$ $y = AJ_0(2\sqrt{kx})\cos nt$, $\quad k = n^2/gm$,

as particular solutions of the equation when (α) the depth is uniform but the breadth is proportional to the distance out to sea, and when (β) the breadth is uniform but the depth is mx. Discuss the shape of the waves that may thus stand on the surface of the estuary or beach.

8. If a series of parallel waves on an ocean of constant depth h is cut perpendicularly by the xy-plane with the axes horizontal and vertical so that $y=-h$ is the ocean bed, the equations for the velocity potential ϕ are known to be

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}=0,\qquad \left[\frac{\partial\phi}{\partial y}\right]_{y=-h}=0,\qquad \left[\frac{\partial^2\phi}{\partial t^2}+g\frac{\partial\phi}{\partial y}\right]_{y=0}=0.$$

Find and combine particular solutions to show that ϕ may have the form

$$\phi=A\cosh k(y+h)\cos(kx-nt),\qquad n^2=gk\tanh kh.$$

9. Obtain the solutions or types of solutions for these equations.

(α) $\dfrac{\partial^2 V}{\partial z^2}+\dfrac{\partial^2 V}{\partial r^2}+\dfrac{1}{r}\dfrac{\partial V}{\partial r}+\dfrac{1}{r^2}\dfrac{\partial^2 V}{\partial\phi^2}=0,$ *Ans.* $e^{\pm kz}\begin{Bmatrix}\cos m\phi\\ \sin m\phi\end{Bmatrix}J_m(kr),$

(β) $\dfrac{\partial^2 V}{\partial r^2}+\dfrac{1}{r}\dfrac{\partial V}{\partial r}+\dfrac{1}{r^2}\dfrac{\partial^2 V}{\partial\phi^2}+V=0,$ *Ans.* $\sum(a_m\cos m\phi+b_m\sin m\phi)J_m(r),$

(γ) $\dfrac{\partial^2 V}{\partial x^2}+\dfrac{\partial^2 V}{\partial y^2}+\dfrac{\partial^2 V}{\partial z^2}+V=0,$ *Ans.* $r^{-\frac{1}{2}}J_{m+\frac{1}{2}}(r)\,P_{n,m}(\cos\theta)\times(a_{n,m}\cos m\phi+b_{n,m}\sin m\phi),$

(δ) $\dfrac{\partial^2 V}{\partial t^2}+2\dfrac{\partial V}{\partial t}=\dfrac{\partial^2 V}{\partial x^2},$ (ϵ) $\dfrac{1}{c^2}\dfrac{\partial^2 V}{\partial t^2}=\dfrac{\partial^2 V}{\partial x^2}+\dfrac{\partial^2 V}{\partial y^2}+\dfrac{\partial^2 V}{\partial z^2}.$

10. Find the potential of a homogeneous circular disk as (Ex. 22, p. 68; Ex. 23, p. 332)

$$V=\frac{2M}{a}\left[\frac{1}{2}\frac{a}{r}-\frac{1\cdot 1}{2\cdot 4}\frac{a^3}{r^3}P_2+\frac{1\cdot 1\cdot 3}{2\cdot 4\cdot 6}\frac{a^5}{r^5}P_4-\frac{1\cdot 1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 8}\frac{a^7}{r^7}P_6+\cdots\right],\qquad r>a,$$

$$=\frac{2M}{a}\left[1\mp\frac{r}{a}P_1+\frac{1}{2}\frac{r^2}{a^2}P_2-\frac{1\cdot 1}{2\cdot 4}\frac{r^4}{a^4}P_4+\frac{1\cdot 1\cdot 3}{2\cdot 4\cdot 6}\frac{r^6}{a^6}P_6-\cdots\right],\qquad r<a,$$

where the negative sign before P_1 holds for $\theta<\frac{1}{2}\pi$ and the positive for $\theta>\frac{1}{2}\pi$.

11. Find the potential of a homogeneous hemispherical shell.

12. Find the potential of (α) a homogeneous hemisphere at all points outside the hemisphere, and (β) a homogeneous circular cylinder at all external points.

13. Assume $\dfrac{Q}{2a}\cos^{-1}\dfrac{x^2-a^2}{x^2+a^2}$ is the potential at a point of the axis of a conducting disk of radius a charged with Q units of electricity. Find the potential anywhere.

196. Harmonic functions; general theorems. A function which satisfies Laplace's equation $V''_{xx}+V''_{yy}=0$ or $V''_{xx}+V''_{yy}+V''_{zz}=0$, whether in the plane or in space, is called a *harmonic function*. It is assumed that the first and second partial derivatives of a harmonic function are continuous except at specified points called singular points. There are many similarities between harmonic functions in the plane and harmonic functions in space, and some differences. The fundamental theorem is that: *If a function is harmonic and has no singularities upon or within a simple closed curve (or surface), the line integral of its normal derivative along the curve (respectively, surface) vanishes; and conversely if a function $V(x, y)$, or $V(x, y, z)$, has continuous first and second*

partial derivatives and the line integral (or surface integral) along every closed curve (or surface) in a region vanishes, the function is harmonic. For by Green's Formula, in the respective cases of plane and space (Ex. 10, p. 349),

$$\int_{\circ} \frac{dV}{dn}\, ds = \int_{\circ} \frac{\partial V}{\partial x}\, dy - \frac{\partial V}{\partial y}\, dx = \iint \left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} \right) dx\, dy,$$

$$\int_{\circ} \frac{dV}{dn}\, dS = \int_{\circ} d\mathbf{S}\cdot\nabla V = \iiint \nabla\cdot\nabla V\, dx\, dy\, dz. \tag{9}$$

Now if the function is harmonic, the right-hand side vanishes and so must the left; and conversely if the left-hand side vanishes for all closed curves (or surfaces), the right-hand side must vanish for every region, and hence the integrand must vanish.

If in particular the curve or surface be taken as a circle or sphere of radius a and polar coördinates be taken at the center, the normal derivative becomes $\partial V/\partial r$ and the result is

$$\int_0^{2\pi} \frac{\partial V}{\partial r}\, d\phi = 0 \quad \text{or} \quad \int_0^{2\pi}\int_0^{\pi} \frac{\partial V}{\partial r} \sin\theta\, d\theta\, d\phi = 0,$$

where the constant a or a^2 has been discarded from the element of arc $a\,d\phi$ or the element of surface $a^2 \sin\theta\, d\theta\, d\phi$. If these equations be integrated with respect to r from 0 to a, the integrals may be evaluated by reversing the order of integration. Thus

$$0 = \int_0^a dr \int_0^{2\pi} \frac{\partial V}{\partial r}\, d\phi = \int_0^{2\pi}\int_0^a \frac{\partial V}{\partial r}\, dr\, d\phi = \int_0^{2\pi} (V_a - V_0)\, d\phi,$$

and $$\int_0^{2\pi} V_a\, d\phi = V_0 \int_0^{2\pi} d\phi, \quad \text{or} \quad \overline{V}_a = V_0, \tag{10}$$

where V_a is the value of V on the circle of radius a and V_0 is the value at the center and $\overline{V}_a$ is the average value along the perimeter of the circle. Similar analysis would hold in space. The result states the important theorem: *The average value of a harmonic function over a circle (or sphere) is equal to the value at the center.*

This theorem has immediate corollaries of importance. *A harmonic function which has no singularities within a region cannot become maximum or minimum at any point within the region.* For if the function were a maximum at any point, that point could be surrounded by a circle or sphere so small that the value of the function at every point of the contour would be less than at the assumed maximum and hence the average value on the contour could not be the value at the center.

A harmonic function which has no singularities within a region and is constant on the boundary is constant throughout the region. For the maximum and minimum values must be on the boundary, and if these have the same value, the function must have that same value throughout the included region. *Two harmonic functions which have identical values upon a closed contour and have no singularities within, are identical throughout the included region.* For their difference is harmonic and has the constant value 0 on the boundary and hence throughout the region. These theorems are equally true if the region is allowed to grow until it is infinite, provided the values which the function takes on at infinity are taken into consideration. Thus, if two harmonic functions have no singularities in a certain infinite region, take on the same values at all points of the boundary of the region, and approach the same values as the point (x, y) or (x, y, z) in any manner recedes indefinitely in the region, the two functions are identical.

If Green's Formula be applied to a product UdV/dn, then

$$\int_{\bigcirc} U\frac{dV}{dn}ds = \int_{\bigcirc} U\frac{dV}{dx}dy - U\frac{dV}{dy}dx$$

$$= \iint U(V''_{xx} + V''_{yy})\,dxdy + \iint (U'_xV'_x + U'_yV'_y)\,dxdy,$$

or
$$\int_{\bigcirc} U d\mathbf{S}\cdot\nabla V = \int U\nabla\cdot\nabla V dv + \int \nabla U\cdot\nabla V dv \tag{11}$$

in the plane or in space. In this relation let V be harmonic without singularities within and upon the contour, and let $U = V$. The first integral on the right vanishes and the second is necessarily positive unless the relations $V'_x = V'_y = 0$ or $V'_x = V'_y = V'_z = 0$, which is equivalent to $\nabla V = 0$, are fulfilled at all points of the included region. Suppose further that the normal derivative dV/dn is zero over the entire boundary. The integral on the left will then vanish and that on the right must vanish. Hence V contains none of the variables and is constant. *If the normal derivative of a function harmonic and devoid of singularities at all points on and within a given contour vanishes identically upon the contour, the function is constant.* As a corollary: If two functions are harmonic and devoid of singularities upon and within a given contour, and if their normal derivatives are identically equal upon the contour, the functions differ at most by an additive constant. In other words, *a harmonic function without singularities not only is determined by its values on a contour but also (except for an additive constant) by the values of its normal derivative upon a contour.*

Laplace's equation arises directly upon the statement of some problems in physics in mathematical form. In the first place consider the flow of heat or of electricity in a conducting body. The physical law is that heat flows along the direction of most rapid decrease of temperature T, and that the amount of the flow is proportional to the rate of decrease. As $-\nabla T$ gives the direction and magnitude of the most rapid decrease of temperature, the flow of heat may be represented by $-k\nabla T$, where k is a constant. The rate of flow in any direction is the component of this vector in that direction. The rate of flow across any boundary is therefore the integral along the boundary of the normal derivative of T. Now the flow is said to be *steady* if there is no increase or decrease of heat within any closed boundary, that is

$$k\int_{\circ} d\mathbf{S}\cdot\nabla T = 0 \quad \text{or} \quad T \text{ is harmonic.}$$

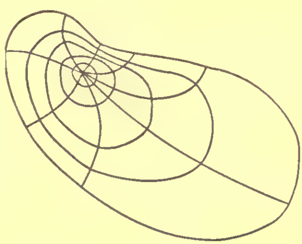
Hence the problem of the distribution of the temperature in a body supporting a steady flow of heat is the problem of integrating Laplace's equation. In like manner, the laws of the flow of electricity being identical with those for the flow of heat except that the potential V replaces the temperature T, the problem of the distribution of potential in a body supporting a steady flow of electricity will also be that of solving Laplace's equation.

Another problem which gives rise to Laplace's equation is that of the irrotational motion of an incompressible fluid. If $\mathbf{v}$ is the velocity of the fluid, the motion is called irrotational when $\nabla\times\mathbf{v} = 0$, that is, when the line integral of the velocity about any closed curve is zero. In this case the negative of the line integral from a fixed limit to a variable limit defines a function $\Phi(x, y, z)$ called the velocity potential, and the velocity may be expressed as $\mathbf{v} = -\nabla\Phi$. As the fluid is incompressible, the flow across any closed boundary is necessarily zero. Hence

$$\int_{C} d\mathbf{S}\cdot\nabla\Phi = 0 \quad \text{or} \quad \int\nabla\cdot\nabla\Phi\, dv = 0 \quad \text{or} \quad \nabla\cdot\nabla\Phi = 0,$$

and the velocity potential Φ is a harmonic function. Both these problems may be stated without vector notation by carrying out the ideas involved with the aid of ordinary coördinates. The problems may also be solved for the plane instead of for space in a precisely analogous manner.

197. The conception of the flow of electricity will be advantageous in discussing the singularities of harmonic functions and a more general conception of steady flow. Suppose an electrode is set down on a sheet of zinc of which the perimeter is grounded. The equipotential lines and the lines of flow which are orthogonal to them may be sketched in. Electricity passes steadily from the electrode to the rim of the sheet and off to the ground. Across any circuit which does not surround the electrode the



flow of electricity is zero as the flow is steady, but across any circuit surrounding the electrode there will be a certain definite flow; the circuit integral of the normal derivative of the potential V around such

a circuit is not zero. This may be compared with the fact that the circuit integral of a function of a complex variable is not necessarily zero about a singularity, although it is zero if the circuit contains no singularity. Or the electrode may not be considered as corresponding to a singularity but to a portion cut out from the sheet so that the sheet is no longer simply connected, and the comparison would then be with a circuit which could not be shrunk to nothing. Concerning this latter interpretation little need be said; the facts are readily seen. It is the former conception which is interesting.

For mathematical purposes the electrode will be idealized by assuming its diameter to shrink down to a point. It is physically clear that the smaller the electrode, the higher must be the potential at the electrode to force a given flow of electricity into the plate. Indeed it may be seen that V must become infinite as $-C\log r$, where r is the distance from the point electrode. For note in the first place that $\log r$ is a solution of Laplace's equation in the plane: and let $U = V + C\log r$ or $V = U - C\log r$, where U is a harmonic function which remains finite at the electrode. The flow across any small circle concentric with the electrode is

$$-\int_0^{2\pi}\frac{\partial V}{\partial r}r\,d\phi = -\int_0^{2\pi}\frac{\partial U}{\partial r}r\,d\phi + 2\pi C = 2\pi C,$$

and is finite. The constant C is called the strength of the source situated at the point electrode. A similar discussion for space would show that the potential in the neighborhood of a source would become infinite as C/r. The particular solutions $-\log r$ and $1/r$ of Laplace's equation in the respective cases may be called the *fundamental solutions.*

The physical analogy will also suggest a method of obtaining higher singularities by combining fundamental singularities. For suppose that a powerful positive electrode is placed near an equally powerful negative electrode, that is, suppose a strong source and a strong sink near together. The greater part of the flow will be nearly in a straight line from the source to the sink, but some part of it will spread out over the sheet. The value of V obtained by adding together the two values for source and sink is

$$\begin{aligned}V &= -\tfrac{1}{2}C\log(r^2+l^2-2rl\cos\phi)+\tfrac{1}{2}C\log(r^2+l^2+2rl\cos\phi)\\ &= -\frac{1}{2}C\log\left(1-\frac{2l}{r}\cos\phi+\frac{l^2}{r^2}\right)+\frac{1}{2}C\log\left(1+\frac{2l}{r}\cos\phi+\frac{l^2}{r^2}\right)\\ &= \frac{2lC}{r}\cos\phi+\text{higher powers}=\frac{M}{r}\cos\phi+\cdots.\end{aligned}$$

Thus if the strength C be allowed to become infinite as the distance $2l$ becomes zero, and if M denote the limit of the product $2lC$, the limiting form of V is $^{1}\cos\phi$ and is itself a solution of the equation, becoming infinite more strongly r. In space the corresponding solution would be $Mr^{-2}\cos\phi$.

It was seen that a harmonic function which had no singularities on or within a given contour was determined by its values on the contour and determined except for an additive constant by the values of its normal derivative upon the contour. If now there be actually within the contour certain singularities at which the function becomes infinite as certain particular solutions $V_1, V_2, \cdots$, the function $U = V - V_1 - V_2 - \cdots$ is harmonic without singularities and may be determined as before. Moreover, the values of $V_1, V_2, \cdots$ or their normal derivatives may be considered as known upon the contour inasmuch as these are definite particular solutions. Hence it appears, as before, that *the harmonic function V is determined by its values on the boundary of the region or (except for an additive constant) by the values of its normal derivative on the boundary, provided the singularities are specified in position and their mode of becoming infinite is given in each case as some particular solution of Laplace's equation.*

Consider again the conducting sheet with its perimeter grounded and with a single electrode of strength unity at some interior point of the sheet. The potential thus set up has the properties that: 1° the potential is zero along the perimeter because the perimeter is grounded; 2° at the position P of the electrode the potential becomes infinite as $-\log r$; and 3° at any other point of the sheet the potential is regular and satisfies Laplace's equation. This particular distribution of potential is denoted by $G(P)$ and is called the Green Function of the sheet relative to P. In space the Green Function of a region would still satisfy 1° and 3°, but in 2° the fundamental solution $-\log r$ would have to be replaced by the corresponding fundamental solution $1/r$. It should be noted that the Green Function is really a function

$$G(P) = G(a, b;\ x, y) \quad \text{or} \quad G(P) = G(a, b, c;\ x, y, z)$$

of four or six variables if the position $P(a, b)$ or $P(a, b, c)$ of the electrode is considered as variable. The function is considered as known only when it is known for any position of P.

If now the symmetrical form of Green's Formula

$$-\iint (u\Delta v - v\Delta u)\,dx\,dy + \int_C \left(u\frac{dv}{dn} - v\frac{du}{dn}\right) ds = 0, \tag{12}$$

where Δ denotes the sum of the second derivatives, be applied to the entire sheet with the exception of a small circle concentric with P and if the choice $u = G$ and $v = V$ be made, then as G and V are harmonic the double integral drops out and

$$\int_C - V\frac{dG}{dn}\,ds - \int_0^{2\pi} G\frac{dV}{dr}\,r\,d\phi + \int_0^{2\pi} V\frac{dG}{dr}\,r\,d\phi = 0. \tag{13}$$

Now let the radius r of the small circle approach 0. Under the assumption that V is devoid of singularities and that G becomes infinite as $-\log r$, the middle integral approaches 0 because its integrand does, and the final integral approaches $2\pi V(P)$. Hence

$$V(P) = \frac{-1}{2\pi}\int_{\bigcirc} V\frac{dG}{dn}ds. \tag{13'}$$

This formula expresses the values of V at any interior point of the sheet in terms of the values of V upon the contour and of the normal derivative of G along the contour. It appears, therefore, that *the determination of the value of a harmonic function devoid of singularities within and upon a contour may be made in terms of the values on the contour provided the Green Function of the region is known.* Hence the particular importance of the problem of determining the Green Function for a given region. This theorem is analogous to Cauchy's Integral (§ 126).

EXERCISES

1. Show that any linear function $ax + by + cz + d = 0$ is harmonic. Find the conditions that a quadratic function be harmonic.

2. Show that the real and imaginary parts of any function of a complex variable are each harmonic functions of (x, y).

3. Why is the sum or difference of any two harmonic functions multiplied by any constants itself harmonic? Is the power of a harmonic function harmonic?

4. Show that the product UV of two harmonic functions is harmonic when and only when $U'_xV'_x + U'_yV'_y = 0$ or $\nabla U \cdot \nabla V = 0$. In this case the two functions are called conjugate or orthogonal. What is the significance of this condition geometrically?

5. Prove the average value theorem for space as for the plane.

6. Show for the plane that if V is harmonic, then

$$U = \int \frac{dV}{dn}ds = \int \frac{\partial V}{\partial x}dy - \frac{\partial V}{\partial y}dx$$

is independent of the path and is the conjugate or orthogonal function to V, and that U is devoid of singularities over any region over which V is devoid of them. Show that $V + iU$ is a function of $z = x + iy$.

7. State the problems of the steady flow of heat or electricity in terms of ordinary coördinates for the case of the plane.

8. Discuss for space the problem of the source, showing that C/r gives a finite flow $4\pi C$, where C is called the strength of the source. Note the presence of the factor 4π in the place of 2π as found in two dimensions.

9. Derive the solution $Mr^{-2}\cos\phi$ for the source-sink combination in space.

10. Discuss the problem of the small magnet or the electric doublet in view of Ex. 9. Note that as the attraction is inversely as the square of the distance, the potential of the force satisfies Laplace's equation in space.

11. Let equal infinite sources and sinks be located alternately at the vertices of an infinitesimal square. Find the corresponding particular solution (α) in the case of the plane, and (β) in the case of space. What combination of magnets does this represent if the point of view of Ex. 10 be taken, and for what purpose is the combination used?

12. Express $V(P)$ in terms of $G(P)$ and the boundary values of V in space.

13. If an analytic function has no singularities within or on a contour, Cauchy's Integral gives the value at any interior point. If there are within the contour certain poles, what must be known in addition to the boundary values to determine the function? Compare with the analogous theorem for harmonic functions.

14. Why were the solutions in § 194 as series the only possible solutions provided they were really solutions? Is there any difficulty in making the same inference relative to the problem of the potential of a circular wire in § 195?

15. Let $G(P)$ and $G(Q)$ be the Green Functions for the same sheet but relative to two different points P and Q. Apply Green's symmetric theorem to the sheet from which two small circles about P and Q have been removed, making the choice $u = G(P)$ and $v = G(Q)$. Hence show that $G(P)$ at Q is equal to $G(Q)$ at P. This may be written as

$$G(a, b;\, x, y) = G(x, y;\, a, b) \quad \text{or} \quad G(a, b, c;\, x, y, z) = G(x, y, z;\, a, b, c).$$

16. Test these functions for the harmonic property, determine the conjugate functions and the allied functions of a complex variable:

(α) xy, $\quad(\beta)$ $x^2y - \frac{1}{3}y^3$, $\quad(\gamma)$ $\frac{1}{2}\log(x^2 + y^2)$,

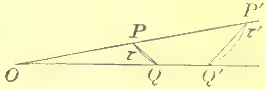
(δ) $e^x \sin x$, $\quad(\epsilon)$ $\sin x \cosh y$, $\quad(\zeta)$ $\tan^{-1}(\cot x \tanh y)$.

198. Harmonic functions; special theorems. For the purposes of the next paragraphs it is necessary to study the properties of the geometric transformation known as *inversion*. The definition of inversion will be given so as to be applicable either to space or to the plane. The transformation which replaces each point P by a point P' such that $OP \cdot OP' = k^2$ where O is a given fixed point, k a constant, and P' is on the line OP, is called *inversion with the center O and the radius k*. Note that if P is thus carried into P', then P' will be carried into P; and hence if any geometrical configuration is carried into another, that other will be carried into the first. Points very near to O are carried off to a great distance; for the point O itself the definition breaks down and O corresponds to no point of space. If desired, one may add to space a fictitious point called the point at infinity and may then say that the center O of the inversion corresponds to the point at infinity (p. 481). A pair of points P, P' which go over into each other, and another pair Q, Q' satisfy the equation $OP \cdot OP' = OQ \cdot OQ'$.

A curve which cuts the line OP at an angle τ is carried into a curve which cuts the line at the angle $\tau' = \pi - \tau$. For by the relation $OP \cdot OP' = OQ \cdot OQ'$, the triangles OPQ, $OQ'P'$ are similar and

$$\angle OPQ = \angle OQ'P' = \pi - \angle O - \angle OP'Q'.$$

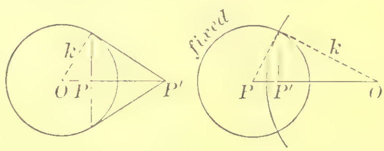
Now if $Q \doteq P$ and $Q' \doteq P'$, then $\angle O \doteq 0$, $\angle OPQ \doteq \tau$, $\angle OP'Q' \doteq \tau$ and it is seen that $\tau = \pi - \tau'$ or $\tau' = \pi - \tau$. An immediate extension of the argument will show that the magnitude of the angle between two intersecting curves will be unchanged by the transformation; *the transformation is therefore conformal.* (In the plane where it is possible to distinguish between positive and negative angles, the sign of the angle is reversed by the transformation.)



If polar coördinates relative to the point O be introduced, the equations of the transformation are simply $rr' = k^2$ with the understanding that the angle ϕ in the plane or the angles ϕ, θ in space are unchanged. The locus $r = k$, which is a circle in the plane or a sphere in space, becomes $r' = k$ and is therefore unchanged. This is called the circle or the sphere of inversion. Relative to this locus a simple construction for a pair of inverse points P and P' may be made as indicated in the figure. The locus

$$r^2 + k^2 = 2\sqrt{a^2 + k^2}\, r \cos\phi \quad \text{becomes} \quad k^2 + r'^2 = 2\sqrt{a^2 + k^2}\, r' \cos\phi$$

and is therefore unchanged as a whole. This locus represents a circle or a sphere of radius a orthogonal to the circle or sphere of inversion. A construction may now be made for finding an inversion which carries a given circle into itself and the center P of the circle into any assigned point P' of the circle; the construction holds for space by revolving the figure about the line OP.



To find what figure a line in the plane or a plane in space becomes on inversion, let the polar axis $\phi = 0$ or $\theta = 0$ be taken perpendicular to the line or plane as the case may be. Then

$$r = p \sec\phi, \qquad r' \sec\phi = k^2/p \quad \text{or} \quad r = p \sec\theta, \qquad r' \sec\theta = k^2/p$$

are the equations of the line or plane and the inverse locus. The locus is seen to be a circle or sphere through the center of inversion. This may also be seen directly by applying the geometric definition of inversion. In a similar manner, or analytically, it may be shown that any circle in the plane or any sphere in space inverts into a circle or into a sphere, unless it passes through the center of inversion and becomes a line or a plane.

If d be the distance of P from the circle or sphere of inversion, the distance of P from the center is $k-d$, the distance of P' from the center is $k^2/(k-d)$, and from the circle or sphere it is $d' = dk/(k-d)$. Now if the radius k is very large in comparison with d, the ratio $k/(k-d)$ is nearly 1 and d' is nearly equal to d. If k is allowed to become infinite so that the center of inversion recedes indefinitely and the circle or sphere of inversion approaches a line or plane, the distance d' approaches d as a limit. As the transformation which replaces each point by a point equidistant from a given line or plane and perpendicularly opposite to the point is the ordinary inversion or reflection in the line or plane such as is familiar in optics, it appears that reflection in a line or plane may be regarded as the limiting case of inversion in a circle or sphere.

The importance of inversion in the study of harmonic functions lies in two theorems applicable respectively to the plane and to space. First, *if V is harmonic over any region of the plane and if that region be inverted in any circle, the function $V'(P') = V(P)$ formed by assigning the same value at P' in the new region as the function had at the point P which inverted into P' is also harmonic.* Second, *if V is harmonic over any region in space, and if that region be inverted in a sphere of radius k, the function $V'(P') = kV(P)/r'$ formed by assigning at P' the value the function had at P multiplied by k and divided by the distance $OP' = r'$ of P' from the center of inversion is also harmonic.* The significance of these theorems lies in the fact that if one distribution of potential is known, another may be derived from it by inversion; and conversely it is often possible to determine a distribution of potential by inverting an unknown case into one that is known. The proof of the theorems consists merely in making the changes of variable

$$r = k^2/r' \quad \text{or} \quad r' = k^2/r, \qquad \phi' = \phi, \qquad \theta' = \theta$$

in the polar forms of Laplace's equation (Exs. 21, 22, p. 112).

The method of using inversion to determine distribution of potential in electrostatics is often called the method of *electric images*. As a charge e located at a point exerts on other point charges a force proportional to the inverse square of the distance, the potential due to e is as $1/\rho$, where ρ is the distance from the charge (with the proper units it may be taken as e/ρ), and satisfies Laplace's equation. The potential due to any number of point charges is the sum of the individual potentials due to the charges. Thus far the theory is essentially the same as if the charges were attracting particles of matter. In electricity, however, the question of the distribution of potential is further complicated when there are in the neighborhood of the charges certain conducting surfaces. For 1° a conducting surface in an electrostatic field must everywhere be at a constant potential or there would be a component force along the surface and the electricity upon it would move, and 2° there is the phenomenon of induced electricity whereby a variable surface charge is induced upon the conductor by other charges in the neighborhood. If the potential $V(P)$ due to any distribution of charges be inverted in any sphere, the new potential is $kV(P)/r'$. As the potential $V(P)$

becomes infinite as e/ρ at the point charges e, the potential $kV(P)/r'$ will become infinite at the inverted positions of the charges. As the ratio $ds':ds$ of the inverted and original elements of length is r'^2/k^2, the potential $kV(P)/r'$ will become infinite as $k/r' \cdot e/\rho' \cdot r'^2/k^2$, that is, as $r'e/k\rho'$. Hence it appears that the charge e inverts into a charge $e' = r'e/k$; the charge $-e'$ is called the electric image of e. As the new potential is $kV(P)/r'$ instead of $V(P)$, it appears that an equipotential surface $V = \text{const.}$ will not invert into an equipotential surface $V'(P') = \text{const.}$ unless $V = 0$ or r' is constant. But if to the inverted system there be added the charge $e = -kV$ at the center O of inversion, the inverted equipotential surface becomes a surface of zero potential.

With these preliminaries, consider the question of the distribution of potential due to an external charge e at a distance r from the center of a conducting spherical surface of radius k which has been grounded so as to be maintained at zero potential. If the system be inverted with respect to the sphere of radius k, the potential of the spherical surface remains zero and the charge e goes over into a charge $e' = r'e/k$ at the inverse point. Now if ρ, ρ' are the distances from e, e' to the sphere, it is a fact of elementary geometry that $\rho : \rho' = \text{const.} = r' : k$. Hence the potential

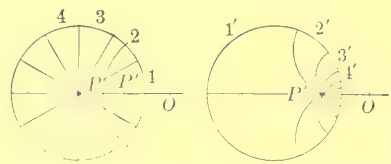
$$V = \frac{e}{\rho} - \frac{e'}{\rho'} = e\left(\frac{1}{\rho} - \frac{r'}{k\rho'}\right) = e\,\frac{k\rho' - r'\rho}{k\rho\rho'},$$

due to the charge e and to its image $-e'$, actually vanishes upon the sphere; and as it is harmonic and has only the singularity e/ρ outside the sphere (which is the same as the singularity due to e), this value of V throughout all space must be precisely the value due to the charge and the grounded sphere. The distribution of potential in the given system is therefore determined. The potential outside the sphere is as if the sphere were removed and the two charges e, $-e'$ left alone. By Gauss's Integral (Ex. 8, p. 348) the charge within any region may be evaluated by a surface integral around the region. This integral over a surface surrounding the sphere is the same as if over a surface shrunk down around the charge $-e'$, and hence the total charge induced on the sphere is $-e' = -r'e/k$.

199. Inversion will transform the average value theorem

$$V(P) = \frac{1}{2\pi}\int_0^{2\pi} V d\phi \quad \text{into} \quad V'(P') = \frac{1}{2\pi}\int_0^{2\pi} V' d\psi, \qquad (14)$$

a form applicable to determine the value of V at any point of a circle in terms of the value upon the circumference. For suppose the circle with center at P and with the set of radii spaced at angles $d\phi$, as implied in the computation of the average value, be inverted upon an orthogonal circle so chosen that P shall go over into P'. The given circle goes over into itself and the series of lines goes over into a series of circles through P' and the center O of inversion. (The figures are drawn separately instead of superposed.) From the conformal property



the angles between the circles of the series are equal to the angles between the radii, and the circles cut the given circle orthogonally just as the radii did Let V' along the arcs $1', 2', 3', \cdots$ be equal to V along the corresponding arcs $1, 2, 3, \cdots$ and let $V(P) = V'(P')$ as required by the theorem on inversion of harmonic functions. Then the two integrals are equal element for element and their values $V(P)$ and $V'(P')$ are equal. Hence the desired form follows from the given form as stated. (It may be observed that $d\phi$ and $d\psi$, strictly speaking, have opposite signs, but in determining the average value $V'(P')$, $d\psi$ is taken positively.) The derived form of integral may be written

$$V'(P') = \frac{1}{2\pi}\int_0^{2\pi} V' d\psi = \frac{1}{2\pi}\int_0^{2\pi a} V' \frac{d\psi}{ds'} ds', \tag{14'}$$

as a line integral along the arc of the circle. If P' is at the distance r from the center, and if a be the radius, the center of inversion O is at the distance a^2/r from the center of the circle, and the value of k is seen to be $k^2 = (a^2 - r^2)a^2/r^2$. Then, if Q and Q' be points on the circle,

$$ds' = ds\frac{\overline{OQ'}^2}{k^2} = \frac{r^2(a^2 - 2a^3r^{-1}\cos\phi' + a^4r^{-2})}{(a^2 - r^2)a^2} a d\phi.$$

Now $d\psi/ds'$ may be obtained, because of the equality of $d\psi$ and $d\phi$, and ds' may be written as $ad\phi'$. Hence

$$V'(P') = \frac{1}{2\pi}\int_0^{2\pi} V'\frac{a^2 - r^2}{a^2 - 2ar\cos\phi' + r^2} d\phi'.$$

Finally the primes may be dropped from V' and P', the position of P' may be expressed in terms of its coördinates (r, ϕ), and

$$V(r, \phi) = \frac{1}{2\pi}\int_0^{2\pi} V\frac{(a^2 - r^2)d\phi'}{a^2 - 2ar\cos(\phi' - \phi) + r^2} = \frac{1}{2\pi}\int_0^{2\pi} V d\psi \tag{15}$$

is the expression of V in terms of its boundary values.

The integral (15) is called *Poisson's Integral.* It should be noted particularly that the form of Poisson's Integral first obtained by inversion represents the average value of V along the circumference, provided that average be computed for each point by considering the values along the circumference as distributed relative to the angle ψ as independent variable. That V as defined by the integral actually approaches the value on the circumference when the point approaches the circumference is clear from the figure, which shows that all except an infinitesimal fraction of the orthogonal circles cut the circle within infinitesimal limits when the point is infinitely near to the circumference. Poisson's Integral may be

obtained in another way. For if P and P' are now two inverse points relative to the circle, the equation of the circle may be written as

$$\rho/\rho' = \text{const.} = r/a, \quad \text{and} \quad G(P) = -\log\rho + \log\rho' + \log(r/a) \tag{16}$$

is then the Green Function of the circular sheet because it vanishes along the circumference, is harmonic owing to the fact that the logarithm of the distance from a point is a solution of Laplace's equation, and becomes infinite at P as $-\log\rho$. Hence

$$V = \frac{-1}{2\pi}\int V\frac{dG}{dn}ds = \frac{-1}{2\pi}\int V\frac{d}{dn}(\log\rho' - \log\rho)\,ds. \tag{16'}$$

It is not difficult to reduce this form of the integral to (15).

If a harmonic function is defined in a region abutting upon a segment of a straight line or an arc of a circle, and if the function vanishes along the segment or arc, the function may be extended across the segment or arc by assigning to the inverse point P' the value $V(P') = -V(P)$, which is the negative of the value at P; the conjugate function

$$U = \int\frac{dV}{dn}ds + C = \int\frac{\partial V}{\partial x}dy - \frac{\partial V}{\partial y}dx + C \tag{17}$$

takes on the same values at P and P'. It will be sufficient to prove this theorem in the case of the straight line because, by the theorem on inversion, the arc may be inverted into a line by taking the center of inversion at any point of the arc or the arc produced. As the Laplace operator $D_x^2 + D_y^2$ is independent of the axes (Ex. 25, p. 112), the line may be taken as the x-axis without restricting the conclusion.

Now the extended function $V(P')$ satisfies Laplace's equation since

$$\frac{\partial^2 V(P')}{\partial x^2} + \frac{\partial^2 V(P')}{\partial y^2} = -\frac{\partial^2 V(P)}{\partial x^2} - \frac{\partial^2 V(P)}{\partial y^2} = 0.$$

Therefore $V(P')$ is harmonic. By the definition $V(P') = -V(P)$ and the assumption that V vanishes along the segment it appears that the function V on the two sides of the line pieces on to itself in a continuous manner, and it remains merely to show that it pieces on to itself in a harmonic manner, that is, that the function V and its extension form a function harmonic at points of the line. This follows from Poisson's Integral applied to a circle centered on the line. For let

$$H(x, y) = \int_0^{2\pi} V d\psi; \quad \text{then} \quad H(x, 0) = 0$$

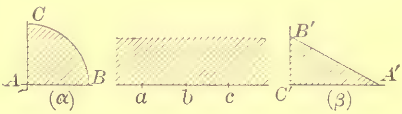
because V takes on equal and opposite values on the upper and lower semicircumferences. Hence $H = V(P) = V(P') = 0$ along the axis. But $H = V(P)$ along the upper arc and $H = V(P')$ along the lower arc because Poisson's Integral takes on the boundary values as a limit when the point approaches the boundary. Now as H is harmonic and agrees with $V(P)$ upon the whole perimeter of the upper semicircle it must be identical with $V(P)$ throughout that semicircle. In like manner

it is identical with $V(P')$ throughout the lower semicircle. As the functions $V(P)$ and $V(P')$ are identical with the single harmonic function H, they must piece together harmonically across the axis. The theorem is thus completely proved. The statement about the conjugate function may be verified by taking the integral along paths symmetric with respect to the axis.

200. *If a function $w = f(z) = u + iv$ of a complex variable becomes real along the segment of a line or the arc of a circle, the function may be extended analytically across the segment or arc by assigning to the inverse point P' the value $w = u - iv$ conjugate to that at P.* This is merely a corollary of the preceding theorem. For if w be real, the harmonic function v vanishes on the line and may be assigned equal and opposite values on the opposite sides of the line; the conjugate function u then takes on equal values on the opposite sides of the line. The case of the circular arc would again follow from inversion as before.

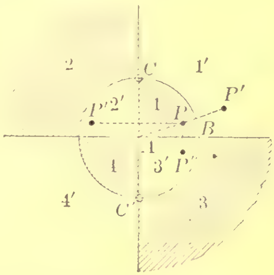
The method employed to identify functions in §§ 185–187 was to map the halves of the w-plane, or rather the several repetitions of these halves which were required to complete the map of the w-surface, on a region of the z-plane. By virtue of the theorem just obtained the converse process may often be carried out and the function $w = f(z)$ which maps a given region of the z-plane upon the half of the w-plane may be obtained. The method will apply only to regions of the z-plane which are bounded by rectilinear segments and circular arcs; for it is only for such that the theorems on inversion and the theorem on the extension of harmonic functions have been proved. To identify the function it is necessary to extend the given region of the z-plane by inversions across its boundaries until the w-surface is completed. The method is not satisfactory if the successive extensions of the region in the z-plane result in overlapping.

The method will be applied to determining the function (α) which maps the first quadrant of the unit circle in the z-plane upon the upper half of the w-plane, and (β) which maps a 30°–60°–90° triangle upon the upper half of the w-plane. Suppose the sector ABC mapped on the w-half-plane so that the perimeter ABC corresponds to the real axis abc. When the perimeter is described in the order written and the interior is on the left, the real axis must, by the principle of conformality, be described in such an order that the upper half-plane which is to correspond to the interior shall also lie on the left. The points a, b, c correspond to points



A, B, C. At these points the correspondence required is such that the conformality must break down. As angles are doubled, each of the points A, B, C must be a critical point of the first order for $w = f(z)$ and a, b, c must be branch points. To map the triangle, similar considerations apply except that whereas C' is a critical point of the first order, the points A', B' are critical of orders 5, 2 respectively. Each case may now be treated separately in detail.

Let it be assumed that the three vertices A, B, C of the sector go into the points* $w = 0, 1, \infty$. As the perimeter of the sector is mapped on the real axis, the function $w = f(z)$ takes on real values for points z along the perimeter. Hence if the sector be inverted over any of its sides, the point P' which corresponds to P may be given a value conjugate to w at P, and the image of P' in the w-plane is symmetrical to the image of P with respect to the real axis. The three regions $1'$, $2'$, $3'$ of the z-plane correspond to the lower half of the w-plane; and the perimeters of these regions correspond also to the real axis. These regions may now be inverted across their boundaries and give rise to the regions 2, 3, 4 which must correspond to the upper half of the w-plane. Finally by inversion from one of these regions the region $4'$ may be obtained as corresponding to the lower half of the w-plane. In this manner the inversion has been carried on until the entire z-plane is covered. Moreover there is no overlapping of the regions and the figure may be inverted in any of its lines without producing any overlapping; it will merely invert into itself. If a Riemann surface were to be constructed over the w-plane, it would clearly require four sheets. The surface could be connected up by studying the correspondence; but this is not necessary. Note merely that the function $f(z)$ becomes infinite at C when $z = i$ by hypothesis and at C' when $z = -i$ by inversion; and at no other point. The values $\pm i$ will therefore be taken as poles of $f(z)$ and as poles of the second order because angles are doubled. Note again that the function $f(z)$ vanishes at A when $z = 0$ by hypothesis and at $z = \infty$ by inversion. These will be assumed to be zeros of the second order because the points are critical points at which angles are doubled. The function



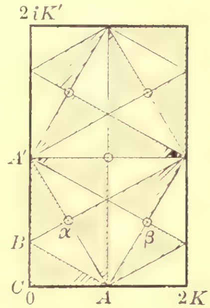
$$w = f(z) = Cz^2(z-i)^{-2}(z+i)^{-2} = Cz^2(z^2+1)^{-2}$$

has the above zeros and poles and must be identical with the desired function when the constant C is properly chosen. As the correspondence is such that $f(1) = 1$ by hypothesis, the constant C is 4. The determination of the function is complete as given.

Consider next the case of the triangle. The same process of inversion and repeated inversion may be followed, and never results in overlapping except as one

* It may be observed that the linear transformation $(\gamma w + \delta)w' = \alpha w + \beta$ (Ex. 15, p. 157) has three arbitrary constants $\alpha : \beta : \gamma : \delta$, and that by such a transformation any three points of the w-plane may be carried into any three points of the w'-plane. It is therefore a proper and trivial restriction to assume that $0, 1, \infty$ are the points of the w-plane which correspond to A, B, C.

region falls into absolute coincidence with one previously obtained. To cover the whole z-plane the inversion would have to be continued indefinitely; but it may be observed that the rectangle inclosed by the heavy line is repeated indefinitely. Hence $w = f(z)$ is a doubly periodic function with the periods $2K$, $2iK'$ if $2K$, $2K'$ be the length and breadth of the rectangle. The function has a pole of the second order at C or $z = 0$ and at the points, marked with circles, into which the origin is carried by the successive inversions. As there are six poles of the second order, the function is of order twelve. When $z = K$ at A or $z = iK'$ at A' the function vanishes and each of these zeros is of the sixth order because angles are increased 6-fold. Again it appears that the function is of order 12. It is very simple to write the function down in terms of the theta functions constructed with the periods $2K$, $2iK'$.



$$w = f(z) = C\frac{H_1^6(z)\,\Theta^6(z)}{H^2(z)\,\Theta_1^2(z)\,H^2(z-\alpha)\,\Theta_1^2(z-\alpha)\,H^2(z-\beta)^2\,\Theta_1^2(z-\beta)}.$$

For this function is really doubly periodic, it vanishes to the sixth order at K, iK', and has poles of the second order at the points

$$0, \quad K + iK', \quad \alpha = \tfrac{1}{2}K + \tfrac{1}{2}iK', \quad \alpha + K + iK', \quad \beta = 2K - \alpha, \quad \beta + K + iK'.$$

As $\beta = 2K - \alpha$ the reduction $H^2(z-\beta) = H^2(z+\alpha)$, $\Theta_1(z-\beta) = \Theta_1(z+\alpha)$ may be made.

$$w = f(z) = C\frac{H_1^6(z)\,\Theta^6(z)}{H^2(z)\,\Theta_1^2(z)\,H^2(z-\alpha)\,H^2(z+\alpha)\,\Theta_1^2(z-\alpha)\,\Theta_1^2(z+\alpha)}.$$

The constant C may be determined, and the expression for $f(z)$ may be reduced further by means of identities; it might be expressed in terms of $\operatorname{sn}(z, k)$ and $\operatorname{cn}(z, k)$, with properly chosen k, or in terms of $p(z)$ and $p'(z)$. For the purposes of computations that might be involved in carrying out the details of the map, it would probably be better to leave the expression of $f(z)$ in terms of the theta functions, as the value of q is about 0.01.

EXERCISES

1. Show geometrically that a plane inverts into a sphere through the center of inversion, and a line into a circle through the center of inversion.

2. Show geometrically or analytically that in the plane a circle inverts into a circle and that in space a sphere inverts into a sphere.

3. Show that in the plane angles are reversed in sign by inversion. Show that in space the magnitude of an angle between two curves is unchanged.

4. If ds, dS, dv are elements of arc, surface, and volume, show that

$$ds' = \frac{r'}{r}\,ds = \frac{r'^2}{k^2}\,ds, \qquad dS' = \frac{r'^2}{r^2}\,dS = \frac{r'^4}{k^4}\,dS, \qquad dv' = \frac{r'^3}{r^3}\,dv = \frac{r'^6}{k^6}\,dv.$$

Note that in the plane an area and its inverted area are of opposite sign, and that the same is true of volumes in space.

5. Show that the system of circles through any point and its inverse with respect to a given circle cut that circle orthogonally. Hence show that if two points are inverse with respect to any circle, they are carried into points inverse with respect to the inverted position of the circle if the circle be inverted in any manner. In particular show that if a circle be inverted with respect to an orthogonal circle, its center is carried into the point which is inverse with respect to the center of inversion.

6. Obtain Poisson's Integral (15) from the form (16′). Note that

$$r^2 = \rho^2 + a^2 - 2a\rho\cos(\rho, n), \qquad \frac{dG}{dn} = \frac{\cos(\rho, n)}{\rho} - \frac{\cos(\rho', n)}{\rho'} = \frac{a^2 - r^2}{a^2\rho^2}.$$

7. From the equation $\rho/\rho' = \text{const.} = r/a$ of the sphere obtain

$$G(P) = \frac{1}{\rho} - \frac{a}{r}\frac{1}{\rho'}, \qquad V = \frac{1}{4\pi a}\int \frac{V(a^2 - r^2)\,dS}{[a^2 + r^2 - 2ar\cos(r, a)]^{\frac{3}{2}}},$$

the Green Function and Poisson's Integral for the sphere.

8. Obtain Poisson's Integral in space by the method of inversion.

9. Find the potential due to an insulated spherical conductor and an external charge (by placing at the center of the sphere a charge equal to the negative of that induced on the grounded sphere).

10. If two spheres intersect at right angles, and charges proportional to the diameters are placed at their centers with an opposite charge proportional to the diameter of the common circle at the center of the circle, then the potential over the two spheres is constant. Hence determine the effect throughout external space of two orthogonal conducting spheres maintained at a given potential.

11. A charge is placed at a distance h from an infinite conducting plane. Determine the potential on the supposition that the plane is insulated with no charge or maintained at zero potential.

12. Map the quadrantal sector on the upper half-plane so that the vertices C, A, B correspond to 1, ∞, 0.

13. Determine the constant C occurring in the map of the triangle on the plane. Find the point into which the median point of the triangle is carried.

14. With various selections of correspondences of the vertices to the three points 0, 1, ∞ of the w-plane, map the following configurations upon the upper half-plane:

(α) a sector of 60°, (β) an isosceles right triangle,

(γ) a sector of 45°, (δ) an equilateral triangle.

201. The potential integrals. If $\rho(x, y, z)$ is a function defined at different points of a region of space, the integral

$$V(\xi, \eta, \zeta) = \iiint \frac{\rho(x, y, z)\,dx\,dy\,dz}{\sqrt{(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2}} = \int \frac{\rho\,dv}{r} \qquad (18)$$

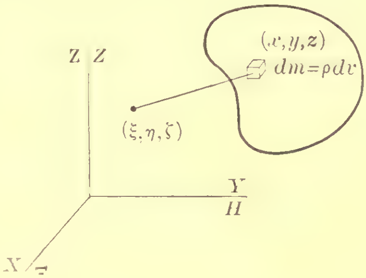
evaluated over that region is called the potential of ρ at the point (ξ, η, ζ). The significance of the integral may be seen by considering the attraction and the potential energy at the point (ξ, η, ζ) due to a

distribution of matter of density $\rho(x, y, z)$ in some region of space. If μ be a mass at (ξ, η, ζ) and m a mass at (x, y, z), the component forces exerted by m upon μ are

$$X = c\frac{\mu m}{r^2}\frac{x-\xi}{r}, \qquad Y = c\frac{\mu m}{r^2}\frac{y-\eta}{r}, \qquad Z = c\frac{\mu m}{r^2}\frac{z-\zeta}{r},$$

and $$F = c\frac{\mu m}{r^2}, \qquad V = -c\mu\frac{m}{r} + C \tag{19}$$

are respectively the total force on μ and the potential energy of the two masses. The potential energy may be considered as the work done by F or X, Y, Z on μ in bringing the mass μ from a fixed point to the point (ξ, η, ζ) under the action of m at (x, y, z) or it may be regarded as the function such that the negative of the derivatives of V by x, y, z give the forces X, Y, Z, or in vector notation $\mathbf{F} = -\nabla V$. Hence if the units be so chosen that $c = 1$, and if the forces and potential at (ξ, η, ζ) be measured per unit mass by dividing by μ, the results are (after disregarding the arbitrary constant C)



$$X = \frac{m}{r^2}\frac{x-\xi}{r}, \qquad Y = \frac{m}{r^2}\frac{y-\eta}{r}, \qquad Z = \frac{m}{r^2}\frac{z-\zeta}{r}, \qquad V = -\frac{m}{r}. \tag{19'}$$

Now if there be a region of matter of density $\rho(x, y, z)$, the forces and potential energy at (ξ, η, ζ) measured per unit mass there located may be obtained by summation or integration and are

$$X = \iiint \frac{\rho(x, y, z)(x-\xi)\,dx\,dy\,dz}{[(\xi-x)^2+(\eta-y)^2+(\zeta-z)^2]^{\frac{3}{2}}}, \ldots, \quad V = -\int\frac{\rho\,dv}{r}. \tag{19''}$$

It therefore appears that the potential U defined by (18) is the negative of the potential energy V due to the distribution of matter.* Note further that in evaluating the integrals to determine X, Y, Z, and $U = -V$, the variables x, y, z with respect to which the integrations are performed will drop out on substituting the limits which determine the region, and will therefore leave X, Y, Z, U as functions of the parameters ξ, η, ζ which appear in the integrand. And finally

$$X = \frac{\partial U}{\partial \xi}, \qquad Y = \frac{\partial U}{\partial \eta}, \qquad Z = \frac{\partial U}{\partial \zeta}. \tag{20}$$

*In electric and magnetic theory, where like *repels* like, the potential and potential energy have the same sign.

are consequences either of differentiating U under the sign of integration or of integrating the expressions (19′) for X, Y, Z expressed in terms of the derivatives of U, over the whole region.

Theorem. The potential integral U satisfies the equations

$$\frac{\partial^2 U}{\partial \xi^2}+\frac{\partial^2 U}{\partial \eta^2}+\frac{\partial^2 U}{\partial \zeta^2}=0 \quad \text{or} \quad \frac{\partial^2 U}{\partial \xi^2}+\frac{\partial^2 U}{\partial \eta^2}+\frac{\partial^2 U}{\partial \zeta^2}=-4\pi\rho, \qquad (21)$$

known respectively as *Laplace's* and *Poisson's Equations*, according as the point (ξ, η, ζ) lies outside or within the body of density $\rho(x, y, z)$.

In case (ξ, η, ζ) lies outside the body, the proof is very simple. For the second derivatives of U may be obtained by differentiating with respect to ξ, η, ζ under the sign of integration, and the sum of the results is then zero. In case (ξ, η, ζ) lies within the body, the value for r vanishes when (ξ, η, ζ) coincides with (x, y, z) during the integration, and hence the integrals for U, X, Y, Z become infinite integrals for which differentiation under the sign is not permissible without justification. Suppose therefore that a small sphere of radius r concentric with (ξ, η, ζ) be cut out of the body, and the contributions $\mathbf{F}'$ of this sphere and $\mathbf{F}^*$ of the remainder of the body to the force $\mathbf{F}$ be considered separately. For convenience suppose the origin moved up to the point (ξ, η, ζ). Then

$$\mathbf{F}=\nabla U=\mathbf{F}^*+\mathbf{F}'=\int_* \rho\nabla\frac{1}{r}\,dv+\mathbf{F}'.$$

Now as the sphere is small and the density ρ is supposed continuous, the attraction F' of the sphere at any point of its surface may be taken as $\frac{4}{3}\pi r^3\rho_0/r^2$, the quotient of the mass by the square of the distance to the center, where ρ_0 is the density at the center. The force $\mathbf{F}'$ then reduces to $-\frac{4}{3}\pi\rho_0\mathbf{r}$ in magnitude and direction. Hence

$$\nabla\cdot\mathbf{F}=\nabla\cdot\nabla U=\nabla\cdot\mathbf{F}^*+\nabla\cdot\mathbf{F}'=\int_* \rho\nabla\cdot\nabla\frac{1}{r}\,dv+\nabla\cdot\mathbf{F}'.$$

The integral vanishes as in the first case, and $\nabla\cdot\mathbf{F}'=-4\pi\rho_0$. Hence if the suffix 0 be now dropped, $\nabla\cdot\nabla U=-4\pi\rho$, and Poisson's Equation is proved. Gauss's Integral (p. 348) affords a similar proof.

A rigorous treatment of the potential U and the forces X, Y, Z and their derivatives requires the discussion of convergence and allied topics. A detailed treatment will not be given, but a few of the most important facts may be pointed out. Consider the ordinary case where the volume density ρ remains finite and the body itself does not extend to infinity. The integrand ρ/r becomes infinite when $r=0$. But as dv is an infinitesimal of the third order around the point where $r=0$, the term $\rho dv/r$ in the integral U will be infinitesimal, may be disregarded, and the integral U converges. In like manner the integrals for X, Y, Z will converge

because $\rho(\xi - x)/r^3$, etc., become infinite at $r = 0$ to only the second order. If $\partial X/\partial\xi$ were obtained by differentiation under the sign, the expressions ρ/r^3 and $\rho(\xi - x)^2/r^5$ would become infinite to the third order, and the integrals

$$\int \frac{\rho}{r^3}\,dv = \iiint \frac{\rho}{r^3}\, r^2 \sin\theta\, dr d\phi d\theta, \text{ etc.,}$$

as expressed in polar coördinates with origin at $r = 0$, are seen to diverge. Hence the derivatives of the forces and the second derivatives of the potential, as obtained by differentiating under the sign, are valueless.

Consider therefore the following device:

$$\frac{\partial}{\partial\xi}\frac{1}{r} = -\frac{\partial}{\partial x}\frac{1}{r}, \qquad \frac{\partial U}{\partial\xi} = \int \rho \frac{\partial}{\partial\xi}\frac{1}{r}\,dv = -\int \rho \frac{\partial}{\partial x}\frac{1}{r}\,dv,$$

$$\frac{\partial}{\partial x}\frac{\rho}{r} = \frac{\partial\rho}{\partial x}\frac{1}{r} + \rho\frac{\partial}{\partial x}\frac{1}{r}, \qquad -\int \rho\frac{\partial}{\partial x}\frac{1}{r}\,dv - \int \frac{1}{r}\frac{\partial\rho}{\partial x}\,dv - \int \frac{\partial}{\partial x}\frac{\rho}{r}\,dv.$$

The last integral may be transformed into a surface integral so that

$$\frac{\partial U}{\partial\xi} = \int \frac{1}{r}\frac{\partial\rho}{\partial x}\,dv - \int \frac{\rho}{r}\cos\alpha\, dS = \iiint \frac{1}{r}\frac{\partial\rho}{\partial x}\,dxdydz - \iint \frac{\rho}{r}\,dydz. \tag{22}$$

It should be remembered, however, that if $r = 0$ within the body, the transformation can only be made after cutting out the singularity $r = 0$, and the surface integral must extend over the surface of the excised region as well as over the surface of the body. But in this case, as dS is of the second order of infinitesimals while r is of the first order, the integral over the surface of the excised region vanishes when $r \doteq 0$ and the equation is valid for the whole region. In vectors

$$\nabla U = \int \frac{\nabla\rho}{r}\,dv - \int \frac{\rho}{r}\,d\mathbf{S}. \tag{22'}$$

It is noteworthy that the first integral gives the potential of $\nabla\rho$, that is, the integral is formed for $\nabla\rho$ just as (18) was from ρ. As $\nabla\rho$ is a vector, the summation is vector addition. It is further noteworthy that in $\nabla\rho$ the differentiation is with respect to x, y, z, whereas in ∇U it is with respect to ξ, η, ζ. Now differentiate (22) under the sign. (Distinguish ∇ as formed for ξ, η, ζ and x, y, z by ∇_ξ and ∇_x.)

$$\frac{\partial^2 U}{\partial\xi^2} = \int \frac{\partial}{\partial\xi}\frac{1}{r}\frac{\partial\rho}{\partial x}\,dv - \int \rho\cos\alpha\frac{\partial}{\partial\xi}\frac{1}{r}\,dS \text{ or } \nabla_\xi\cdot\nabla_\xi U = \int \nabla_\xi\frac{1}{r}\cdot\nabla_x\rho\, dv - \int \rho\nabla_\xi\frac{1}{r}\cdot d\mathbf{S},$$

or again
$$\nabla_\xi\cdot\nabla_\xi U = -\int \nabla_x\frac{1}{r}\cdot\nabla_x\rho\, dv + \int \rho\nabla_x\frac{1}{r}\cdot d\mathbf{S}. \tag{23}$$

This result is valid for the whole region. Now by Green's Formula (Ex. 10, p. 349)

$$\int \rho\nabla_x\cdot\nabla_x\frac{1}{r}\,dv + \int \nabla_x\frac{1}{r}\cdot\nabla_x\rho\, dv = \int \nabla_x\cdot\left(\rho\nabla_x\frac{1}{r}\right)dv = \int \rho\nabla_x\frac{1}{r}\cdot d\mathbf{S} = \int \rho\frac{d}{dn}\frac{1}{r}\,dS.$$

Here the small region about $r = 0$ must again be excised and the surface integral must extend over its surface. If the region be taken as a sphere, the normal dn, being exterior to the body, is directed along $-dr$. Thus for the sphere

$$\int \rho\frac{d}{dn}\frac{1}{r}\,dS = \iint \rho\frac{1}{r^2}\, r^2\sin\theta\, d\phi d\theta = \iint \rho\sin\theta\, d\phi d\theta = 4\pi\rho,$$

where $\bar{\rho}$ is the average of ρ upon the surface. If now r be allowed to approach 0 and $\nabla\cdot\nabla r^{-1}$ be set equal to zero, Green's Formula reduces to

$$\int \nabla_x \frac{1}{r} \cdot \nabla_x \rho dv = \int \rho \nabla_x \frac{1}{r} \cdot d\mathbf{S} + 4\pi\rho,$$

where the volume integrals extend over the whole volume and the surface integral extends like that of (23) over the surface of the body but not over the small sphere. Hence (23) reduces to $\nabla\cdot\nabla U = -4\pi\rho$.

Throughout this discussion it has been assumed that ρ and its derivatives are continuous throughout the body. In practice it frequently happens that a body consists really of several, say two, bodies of different nature (separated by a bounding surface S_{12}) in each of which ρ and its derivatives are continuous. Let the suffixes 1, 2 serve to distinguish the bodies. Then

$$U = \int \frac{\rho_1}{r} dv_1 + \int \frac{\rho_2}{r} dv_2 = \int \frac{\rho}{r} dv.$$

The discontinuity in ρ along a surface S_{12} does not affect a triple integral.

$$\nabla U = \int \frac{\nabla \rho_1}{r} dv_1 - \int \frac{\rho_1}{r} d\mathbf{S}_{1,12} + \int \frac{\nabla \rho_2}{r} dv_2 - \int \frac{\rho_2}{r} d\mathbf{S}_{2,21}.$$

Here the first surface integral extends over the boundary of the region 1 which includes the surface S_{12} between the regions. For the interface S_{12} the direction of $d\mathbf{S}$ is from 1 into 2 in the first case, but from 2 into 1 in the second. Hence

$$\nabla U = \int \frac{\nabla \rho}{r} dv - \int \frac{\rho}{r} d\mathbf{S} - \int \frac{\rho_1 - \rho_2}{r} d\mathbf{S}_{12}.$$

It may be noted that the first and second surface integrals are entirely analogous because the first may be regarded as extended over the surface separating a body of density ρ from one of density 0. Now $\nabla\cdot\nabla U$ may be found, and if the proper modifications be introduced in Green's Formula, it is seen that $\nabla\cdot\nabla U = -4\pi\rho$ still holds provided the point lies entirely within either body. The fact that ρ comes from the average value $\bar{\rho}$ upon the surface of an infinitesimal sphere shows that if the point lies on the interface S_{12} at a regular point, $\nabla\cdot\nabla U = -4\pi(\frac{1}{2}\rho_1 + \frac{1}{2}\rho_2)$.

The application of Green's Formula in its symmetric form (Ex. 10, p. 349) to the two functions r^{-1} and U, and the calculation of the integral over the infinitesimal sphere about $r = 0$, gives

$$\int \left(\frac{1}{r} \nabla\cdot\nabla U - U \nabla\cdot\nabla \frac{1}{r}\right) dv - \int \left(\frac{1}{r}\frac{dU}{dn} - U \frac{d}{dn}\frac{1}{r}\right) dS - 4\pi U$$

or

$$\int \frac{\nabla\cdot\nabla U}{r} dv = \sum \int \frac{\left(\frac{dU}{dn}\right)_1 - \left(\frac{dU}{dn}\right)_2}{r} dS_{12}$$
$$- \sum \int (U_1 - U_2) \frac{d}{dn}\frac{1}{r} dS_{12} - 4\pi U, \qquad (24)$$

where Σ extends over all the surfaces of discontinuity, including the boundary of the whole body where the density changes to 0. Now $\nabla\cdot\nabla U = -4\pi\rho$ and if the definitions be given that

$$\left(\frac{dU}{dn}\right)_1 - \left(\frac{dU}{dn}\right)_2 = -4\pi\sigma, \qquad U_1 - U_2 = 4\pi\tau,$$

then $$U = \int \frac{\rho}{r} dv + \int \frac{\sigma}{r} dS + \int \tau \frac{d}{dn} \frac{1}{r} dS, \tag{25}$$

where the surface integrals extend over all surfaces of discontinuity. This form of U appears more general than the initial form (18), and indeed it is more general, for it takes into account the discontinuities of U and its derivative, which cannot arise when ρ is an ordinary continuous function representing a volume distribution of matter. The two surface integrals may be interpreted as due to surface distributions. For suppose that along some surface there is a surface density σ of matter. Then the first surface integral represents the potential of the matter in the surface. Strictly speaking, a surface distribution of matter with σ units of matter per unit surface is a physical impossibility, but it is none the less a convenient mathematical fiction when dealing with thin sheets of matter or with the charge of electricity upon a conducting surface. The surface distribution may be regarded as a limiting case of volume distribution where ρ becomes infinite and the volume throughout which it is spread becomes infinitely thin. In fact if dn be the thickness of the sheet of matter $\rho dn dS = \sigma dS$. The second surface integral may likewise be regarded as a limit. For suppose that there are two surfaces infinitely near together upon one of which there is a surface density $-\sigma$, and upon the other a surface density σ. The potential due to the two equal superimposed elements dS is the

$$\frac{\sigma_1 dS_1}{r_1} + \frac{\sigma_2 dS_2}{r_2} = \sigma dS \left(\frac{1}{r_2} - \frac{1}{r_1} \right) = \sigma dS \frac{d}{dn} \frac{1}{r} \cdot dn = \sigma dn \frac{d}{dn} \frac{1}{r} dS.$$

Hence if $\sigma dn = \tau$, the potential takes the form $\tau d r^{-1}/dn dS$. Just this sort of distribution of magnetism arises in the case of a magnetic shell, that is, a surface covered on one side with positive poles and on the other with negative poles. The three integrals in (25) are known respectively as volume potential, surface potential, and double surface potential.

202. The potentials may be used to obtain particular integrals of some differential equations. In the first place the equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = f(x, y, z) \quad \text{has} \quad U = \frac{-1}{4\pi} \int \frac{f dv}{r}$$

as its solution, when the integral is extended over the region throughout which f is defined. To this particular solution for U may be added any solution of Laplace's equation, but the particular solution is frequently precisely that particular solution which is desired. If the functions $\mathbf{U}$ and $\mathbf{f}$ were vector functions so that $\mathbf{U} = \mathbf{i}U_1 + \mathbf{j}U_2 + \mathbf{k}U_3$, and $\mathbf{f} = \mathbf{i}f_1 + \mathbf{j}f_2 + \mathbf{k}f_3$, the results would be

$$\frac{\partial^2 \mathbf{U}}{\partial x^2} + \frac{\partial^2 \mathbf{U}}{\partial y^2} + \frac{\partial^2 \mathbf{U}}{\partial z^2} = \mathbf{f}(x, y, z) \quad \text{and} \quad \mathbf{U} = \frac{-1}{4\pi} \int \frac{\mathbf{f} dv}{r},$$

where the integration denotes vector summation, as may be seen by adding the results for $\nabla \cdot \nabla U_1 = f_1$, $\nabla \cdot \nabla U_2 = f_2$, $\nabla \cdot \nabla U_3 = f_3$ after multiplication by $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. If it is desired to indicate the vectorial nature of $\mathbf{U}$ and $\mathbf{f}$, the potential $\mathbf{U}$ may be called a vector potential.

In evaluating the potential and the forces at (ξ, η, ζ) due to an element dm at (x, y, z), it has been assumed that the action depends solely on the distance r. Now suppose that the distribution $\rho(x, y, z, t)$ is a function of the time and that the action of the element ρdv at (x, y, z) does not make its effect felt instantly at (ξ, η, ζ) but is propagated toward (ξ, η, ζ) from (x, y, z) at a velocity $1/a$ so as to arrive at the time $(t + ar)$. The potential and the forces at (ξ, η, ζ) as calculated by (18) will then be those there transpiring at the time $t + ar$ instead of at the time t. To obtain the effect at the time t it would therefore be necessary to calculate the potential from the distribution $\rho(x, y, z, t - ar)$ at the time $t - ar$. The potential

$$\begin{aligned} U(x, y, z, t) &= \int \frac{\rho(x, y, z, t - ar)\, dxdydz}{\sqrt{(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2}} \\ &= \int \frac{\rho(t)}{r} dv + \int \frac{\rho(t - ar) - \rho(t)}{r} dv, \end{aligned} \tag{26}$$

where for brevity the variables x, y, z have been dropped in the second form, is called a *retarded potential* as the time has been set back from t to $t - ar$. *The retarded potential satisfies the equation*

$$\frac{\partial^2 U}{\partial \xi^2} + \frac{\partial^2 U}{\partial \eta^2} + \frac{\partial^2 U}{\partial \zeta^2} - a^2 \frac{\partial^2 U}{\partial t^2} = -4\pi\rho(\xi, \eta, \zeta, t) \quad \text{or} \quad 0 \tag{27}$$

according as (ξ, η, ζ) *lies within or outside the distribution* ρ. There is really no need of the alternative statements because if (ξ, η, ζ) is outside, ρ vanishes. Hence a solution of the equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} - a^2 \frac{\partial^2 U}{\partial t^2} = f(x, y, z, t)$$

is

$$U = \frac{-1}{4\pi} \int \frac{f(x, y, z, t - ar)}{r} dv.$$

The proof of the equation (27) is relatively simple. For in vector notation,

$$\begin{aligned} \nabla\cdot\nabla U &= \nabla\cdot\nabla \int \frac{\rho(t)}{r} dv + \nabla\cdot\nabla \int \frac{\rho(t - ar) - \rho(t)}{r} dv \\ &= -4\pi\rho + \nabla\cdot\nabla \int \frac{\rho(t - ar) - \rho(t)}{r} dv. \end{aligned}$$

The first reduction is made by Poisson's Equation. The second expression may be evaluated by differentiation under the sign. For it should be remarked that $\rho(t - ar) - \rho(t)$ vanishes when $r = 0$, and hence the order of the infinite in the integrand before and after differentiation is less by unity than it was in the corresponding steps of § 201. Then

$$\nabla_\xi \int \frac{\rho(t - ar) - \rho(t)}{r} dv = \int \left\{ \frac{(-a)\rho'(t - ar)\nabla_\xi r}{r} + [\rho(t - ar) - \rho(t)]\nabla_\xi \frac{1}{r} \right\} dv,$$

$$\nabla_\xi\cdot\nabla_\xi \int \frac{\rho(t-ar)-\rho(t)}{r}dv = \int \left\{ \frac{(-a)^2\rho''\nabla_\xi r\cdot\nabla_\xi r}{r} + \frac{(-a)\rho'\nabla_\xi\cdot\nabla_\xi r}{r} \right.$$

$$\left. + (-a)\rho'\nabla_\xi r\cdot\nabla_\xi\frac{1}{r} + (-a)\rho'\nabla_\xi r\cdot\nabla_\xi\frac{1}{r} + [\rho(t-ar)-\rho(t)]\nabla_\xi\cdot\nabla_\xi\frac{1}{r}\right\}dv.$$

But $\quad \nabla_\xi = -\nabla_x \quad$ and $\quad \nabla r = \mathbf{r}/r$ and $\nabla r^{-1} = -\mathbf{r}/r^3 \quad$ and $\quad \nabla\cdot\nabla r^{-1} = 0.$

Hence $\qquad \nabla_\xi r\cdot\nabla_\xi r = 1, \qquad \nabla_\xi r\cdot\nabla_\xi r^{-1} = -r^{-2}, \qquad \nabla_\xi\cdot\nabla_\xi r = 2r^{-1}$

and $\quad \nabla\cdot\nabla\displaystyle\int \frac{\rho(t-ar)-\rho(t)}{r}dv = \int\frac{a^2\rho''}{r}dv = \int\frac{a^2}{r}\frac{\partial^2\rho(t-ar)}{\partial t^2}dv = a^2\frac{\partial^2 U}{\partial t^2}.$

It was seen (p. 345) that if $\mathbf{F}$ is a vector function with no curl, that is, if $\nabla\times\mathbf{F} = 0$, then $\mathbf{F}\cdot d\mathbf{r}$ is an exact differential $d\phi$; and $\mathbf{F}$ may be expressed as the gradient of ϕ, that is, as $\mathbf{F} = \nabla\phi$. This problem may also be solved by potentials. For suppose

$$\mathbf{F} = \nabla\phi, \quad \text{then} \quad \nabla\cdot\mathbf{F} = \nabla\cdot\nabla\phi, \qquad \phi = \frac{-1}{4\pi}\int\frac{\nabla\cdot\mathbf{F}}{r}dv. \tag{28}$$

It appears therefore that ϕ may be expressed as a potential. This solution for ϕ is less general than the former because it depends on the fact that the potential integral of $\nabla\cdot\mathbf{F}$ shall converge. Moreover as the value of ϕ thus found is only a particular solution of $\nabla\cdot\mathbf{F} = \nabla\cdot\nabla\phi$, it should be proved that for this ϕ the relation $\mathbf{F} = \nabla\phi$ is actually satisfied. The proof will be given below. A similar method may now be employed to show that if $\mathbf{F}$ is a vector function with no divergence, that is, if $\nabla\cdot\mathbf{F} = 0$, then $\mathbf{F}$ may be written as the curl of a vector function $\mathbf{G}$, that is, as $\mathbf{F} = \nabla\times\mathbf{G}$. For suppose

$$\mathbf{F} = \nabla\times\mathbf{G}, \quad \text{then} \quad \nabla\times\mathbf{F} = \nabla\times\nabla\times\mathbf{G} = \nabla\nabla\cdot\mathbf{G} - \nabla\cdot\nabla\mathbf{G}.$$

As $\mathbf{G}$ is to be determined, let it be supposed that $\nabla\cdot\mathbf{G} = 0$.

$$\text{Then} \qquad \mathbf{F} = \nabla\times\mathbf{G} \quad \text{gives} \quad \mathbf{G} = \frac{1}{4\pi}\int\frac{\nabla\times\mathbf{F}}{r}dv. \tag{29}$$

Here again the solution is valid only when the vector potential integral of $\nabla\times\mathbf{F}$ converges, and it is further necessary to show that $\mathbf{F} = \nabla\times\mathbf{G}$. The conditions of convergence are, however, satisfied for the functions that usually arise in physics.

To amplify the treatment of (28) and (29), let it be shown that

$$\nabla\phi = -\frac{1}{4\pi}\nabla\int\frac{\nabla\cdot\mathbf{F}}{r}dv = \mathbf{F}, \qquad \nabla\times\mathbf{G} = \frac{1}{4\pi}\nabla\times\int\frac{\nabla\times\mathbf{F}}{r}dv = \mathbf{F}.$$

By use of (22) it is possible to pass the differentiations under the sign of integration and apply them to the functions $\nabla\cdot\mathbf{F}$ and $\nabla\times\mathbf{F}$, instead of to $1/r$ as would be required by Leibniz's Rule (§ 119). Then

$$\nabla\phi = -\frac{1}{4\pi}\int\frac{\nabla\nabla\cdot\mathbf{F}}{r}dv + \frac{1}{4\pi}\int\frac{\nabla\cdot\mathbf{F}}{r}d\mathbf{S}$$

The surface integral extends over the surfaces of discontinuity of $\nabla\cdot\mathbf{F}$, over a large (infinite) surface, and over an infinitesimal sphere surrounding $r = 0$. It will be assumed that $\nabla\cdot\mathbf{F}$ is such that the surface integral is infinitesimal. Now as $\nabla\times\mathbf{F} = 0$, $\nabla\times\nabla\times\mathbf{F} = 0$ and $\nabla\nabla\cdot\mathbf{F} = \nabla\cdot\nabla\mathbf{F}$. Hence if $\mathbf{F}$ and its derivatives are continuous, a reference to (24) shows that

$$\nabla\phi = -\frac{1}{4\pi}\int\frac{\nabla\cdot\nabla\mathbf{F}}{r}\,dv = \mathbf{F}.$$

In like manner

$$\nabla\times\mathbf{G} = \frac{1}{4\pi}\int\frac{\nabla\times\nabla\times\mathbf{F}}{r}\,dv - \frac{1}{4\pi}\int\frac{\nabla\times\mathbf{F}}{r}\times d\mathbf{S} = \frac{-1}{4\pi}\int\frac{\nabla\cdot\nabla\mathbf{F}}{r}\,dv = \mathbf{F}.$$

Questions of continuity and the significance of the vanishing of the neglected surface integrals will not be further examined. The elementary facts concerning potentials are necessary knowledge for students of physics (especially electromagnetism); the detailed discussion of the subject, whether from its physical or mathematical side, may well be left to special treatises.

EXERCISES

1. Discuss the potential U and its derivative ∇U for the case of a uniform sphere, both at external and internal points, and upon the surface.

2. Discuss the second derivatives of the potential, that is, the derivatives of the forces, at a surface of discontinuity of density.

3. If a distribution of matter is external to a sphere, the average value of the potential on the spherical surface is the value at the center; if it is internal, the average value is the value obtained by concentrating all the mass at the center.

4. What density of distribution is indicated by the potential e^{-r^2}? What density of distribution gives a potential proportional to itself?

5. In a space free of matter the determination of a potential which shall take assigned values on the boundary is equivalent to the problem of minimizing

$$\frac{1}{2}\iiint\left[\left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial z}\right)^2\right]dxdydz = \frac{1}{2}\int\nabla U\cdot\nabla U\,dv.$$

6. For Laplace's equation in the plane and for the logarithmic potential $-\log r$, develop the theory of potential integrals analogously to the work of § 201 for Laplace's equation in space and for the fundamental solution $1/r$.

BOOK LIST

A short list of typical books with brief comments is given to aid the student of this text in selecting material for collateral reading or for more advanced study.

1. Some standard elementary differential and integral calculus.

For reference the book with which the student is familiar is probably preferable. It may be added that if the student has had the misfortune to take his calculus under a teacher who has not led him to acquire an easy formal knowledge of the subject, he will save a great deal of time in the long run if he makes up the deficiency soon and thoroughly; practice on the exercises in Granville's Calculus (Ginn and Company), or Osborne's Calculus (Heath & Co.), is especially recommended.

2. B. O. Peirce, *Table of Integrals* (new edition). Ginn and Company.

This table is frequently cited in the text and is well-nigh indispensable to the student for constant reference.

3. Jahnke-Emde, *Funktionentafeln mit Formeln und Kurven.* Teubner.

A very useful table for any one who has numerical results to obtain from the analysis of advanced calculus. There is very little duplication between this table and the previous one.

4. Woods and Bailey, *Course in Mathematics.* Ginn and Company.

5. Byerly, *Differential Calculus* and *Integral Calculus.* Ginn and Company.

6 Todhunter, *Differential Calculus* and *Integral Calculus.* Macmillan.

7. Williamson, *Differential Calculus* and *Integral Calculus.* Longmans.

These are standard works in two volumes on elementary and advanced calculus. As sources for additional problems and for comparison with the methods of the text they will prove useful for reference.

8. C. J. de la Vallée-Poussin, *Cours d' analyse.* Gauthier-Villars.

There are a few books which inspire a positive affection for their style and beauty in addition to respect for their contents, and this is one of those few. My Advanced Calculus is necessarily under considerable obligation to de la Vallée-Poussin's Cours d' analyse, because I taught the subject out of that book for several years and esteem the work more highly than any of its compeers in any language.

9. Goursat, *Cours d'analyse.* Gauthier-Villars.

10. Goursat-Hedrick, *Mathematical Analysis.* Ginn and Company.

The latter is a translation of the first of the two volumes of the former. These, like the preceding five works, will be useful for collateral reading.

11. Bertrand, *Calcul différentiel* and *Calcul intégral.*

This older French work marks in a certain sense the acme of calculus as a means of obtaining formal and numerical results. Methods of calculation are not now so prominent, and methods of the theory of functions are coming more to the fore. Whether this tendency lasts or does not, Bertrand's Calculus will remain an inspiration to all who consult it.

12. Forsyth, *Treatise on Differential Equations.* Macmillan.

As a text on the solution of differential equations Forsyth's is probably the best. It may be used for work complementary and supplementary to Chapters VIII–X of this text.

13. Pierpont, *Theory of Functions of Real Variables.* Ginn and Company.

In some parts very advanced and difficult, but in others quite elementary and readable, this work on rigorous analysis will be found useful in connection with Chapter II and other theoretical portions of our text.

14. Gibbs-Wilson, *Vector Analysis.* Scribners.

Herein will be found a detailed and connected treatment of vector methods mentioned here and there in this text and of fundamental importance to the mathematical physicist.

15. B. O. Peirce, *Newtonian Potential Function.* Ginn and Company.

A text on the use of the potential in a wide range of physical problems. Like the following two works, it is adapted, and practically indispensable, to all who study higher mathematics for the use they may make of it in practical problems.

16. Byerly, *Fourier Series and Spherical Harmonics.* Ginn and Company.

Of international repute, this book presents the methods of analysis employed in the solution of the differential equations of physics. Like the foregoing, it gives an extended development of some questions briefly treated in our Chapter XX.

17. Whittaker, *Modern Analysis.* Cambridge University Press.

This is probably the only book in any language which develops and applies the methods of the theory of functions for the purpose of deriving and studying the formal properties of the most important functions other than elementary which occur in analysis directed toward the needs of the applied mathematician.

18. Osgood, *Lehrbuch der Funktionentheorie.* Teubner.

For the pure mathematician this work, written with a grace comparable only to that of de la Vallée-Poussin's Calculus, will be as useful as it is charming.

INDEX

(The numbers refer to pages)